

# Instructions for Open Source Object Oriented Classification

A tutorial on how to use open-source software, the ORFEO application in QGIS and remotely-sensed satellite data to perform object-oriented classifications.

By: Miguel A. Pavon, MAP

for

Desert Landscape Conservation Cooperative

For help or information on classifications, OTB or QGIS, please keep the online help as reference:

- <https://www.orfeo-toolbox.org/documentation/>
- <https://www.qgis.org/en/site/forusers/index.html>

Additional help may be provided by contacting Miguel Pavon at pavonma@hotmail.com or (512) 466-3936.

Suggested citation: Pavon, Miguel A., Instructions for Open Source Object Oriented Classification, 2017.

## Table of contents.

1. Download and install QGIS software for your platform	2
1.1 Download QGIS Software	2
1.2 Install QGIS Software	3
1.3 Verify the installation of your QGIS Software	10
2. Earth Explorer select and download an image to use in your Object-Oriented Classification	13
2.1 Landsat 8	13
2.2 Create your login at Earth Explorer	13
2.3 Search and select the data you need	14
2.4 Download your data	17
3. Object Oriented Classification with QGIS with ORFEO Toolbox	21
3.1 Sample polygons and ground-truthing	21
3.2 Create the image stack	23
3.3 Vector segmentation and data exploration	27
3.4 Raster Segmentation and Object-Oriented Classification	34
3.4.1 Mean-shift segmentation workflow	34
3.4.2 Preparation of reference data	39
3.4.3 Object-based classification using SVM algorithm	40
3.4.4 Accuracy Assessment	47

## Instructions for Open Source Object Oriented Classification

This set of instructions was compiled by Miguel A. Pavon, MAP, working with Bird Conservancy of the Rockies for the Desert Landscape Conservation Cooperative (DLCC). The rationale behind it is to make available a tutorial for land cover classification that is based on easily accessible data and tools that everybody can use to improve conservation, planning and land management practices. Object-oriented image classification involves identification of image objects, or segments, which are composed of neighboring (spatially contiguous) pixels of similar spectral content (color). The case in hand uses open-source QGIS software, the ORFEO application in QGIS and remotely-sensed, satellite data from Landsat made available by USGS via Earth Explorer.

This tutorial is broken into three sections, the first is the installation of QGIS software, second is the data download (Landsat) from Earth Explorer and lastly instructions on how to use ORFEO within QGIS to perform an object-oriented classification on the Landsat image.

QGIS is a Geographic Information System that runs in multiple platforms, making it ideal for new users and stakeholders of the DLCC to use within the region. It is user-friendly and well-documented, although it has a great variety of tools which can make it look challenging, but the more time you spend using it, the more intuitive it gets.

Earth Explorer is a website-based system to query and order satellite images, aerial photographs and other geographic and cartographic products through the U.S. Geological Survey. The user can specify a search criteria, select datasets, and browse the results before downloading the image(s) and product(s) needed for their research.

ORFEO stands for Optical and Radar Federated Earth Observation and is commonly downloaded as a toolbox that can be used within QGIS. One of the advantages is that it allows object based image analysis and classification.

We thank DLCC for making funds accessible for this project and facilitating webinars, Bird Conservancy of the Rockies for managing the project, QGIS and ORFEO teams for their collaboration on making open-source software integrated and available and USGS for supporting data collection and dissemination on the Earth Explorer User Interface.

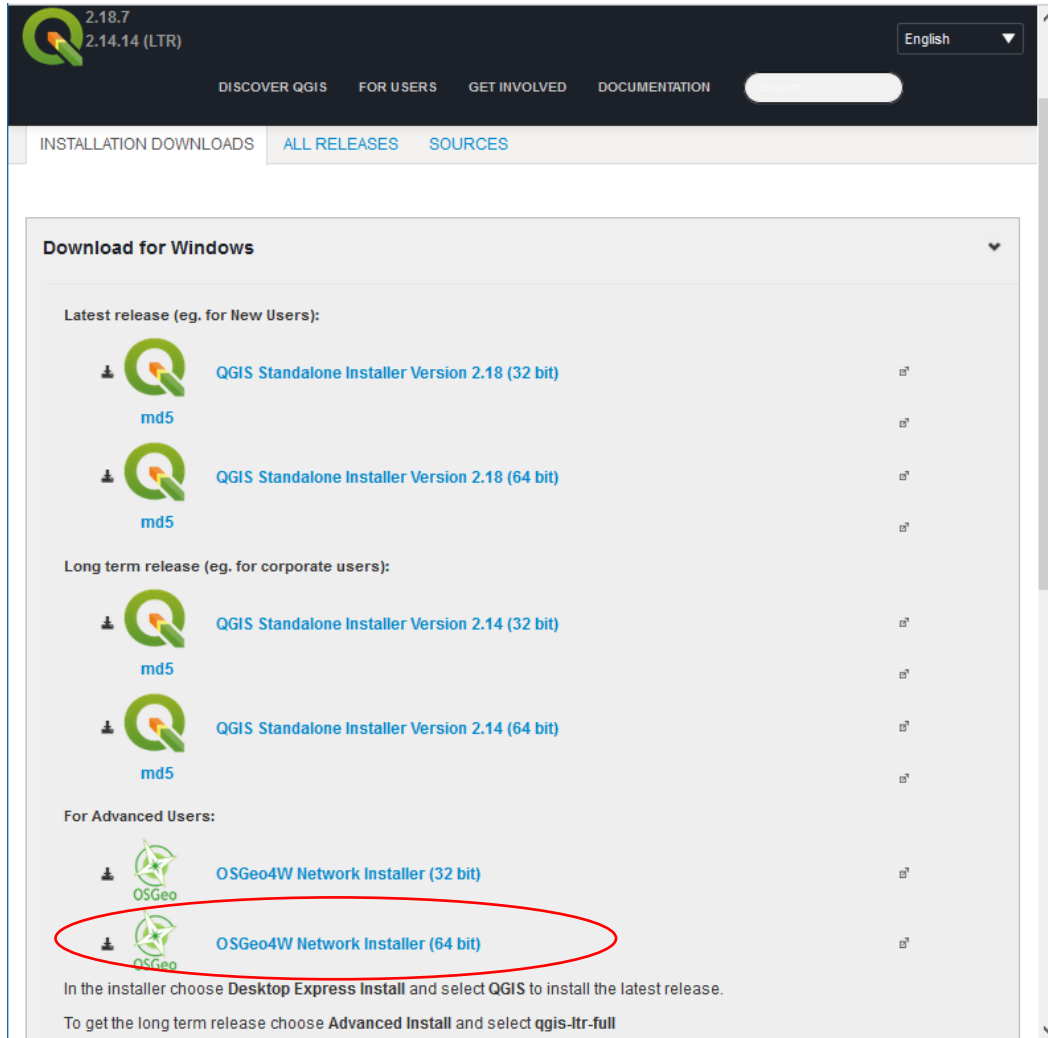
## 1. Download and install QGIS software for your platform

### 1.1 Download QGIS Software

Open your browser and navigate to: <https://www.qgis.org/en/site/forusers/download.html> or <https://www.qgis.org/es/site/forusers/download.html> (en Español)

You will notice the current version, date of release, and that QGIS is available on Windows, MacOS X, Linux and Android in several languages. In this example, we will work with a Windows 64 bit installation,

but you can pick and choose what is most appropriate for your needs. There are two options: the latest release (newer tools, but also the “opportunity” to discover bugs) or the long term release, which is more stable. Instead of taking the more conservative path with the long-term release for 64 bit, we will select the advanced users option: **OSGeo4W Network Installer (64 bit)** near the bottom of the Downloads for Windows that includes the Orfeo Tool Box (OTB).



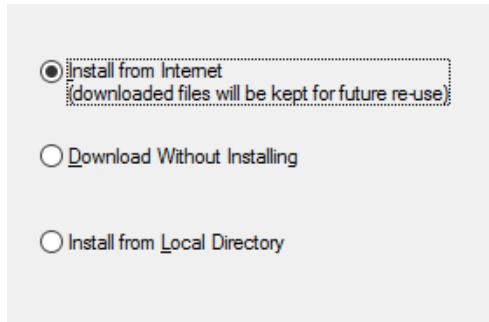
Download the software and start the installation.

### 1.2 Install QGIS Software

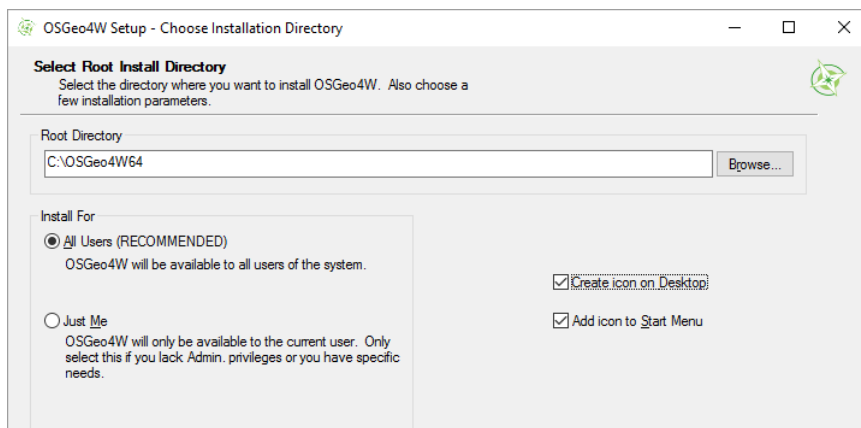
Click “**Advanced Install**” and then **Next** to continue.

Choose to install from Internet (each component will be chosen and then downloaded individually from the repository). Click Next.

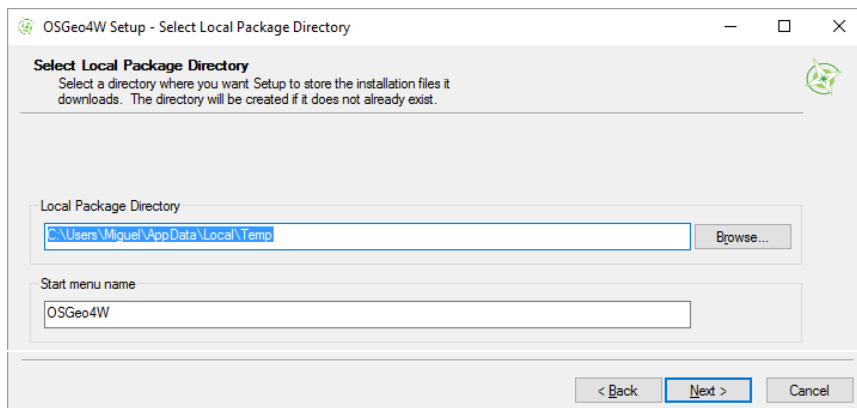
## Instructions for Open Source Object Oriented Classification



Select the root directory for the installation, and choose the option to install the software with availability for all users. If you want to create an icon on the Desktop and to add it to the Start menu, choose this option as well, click **Next**.

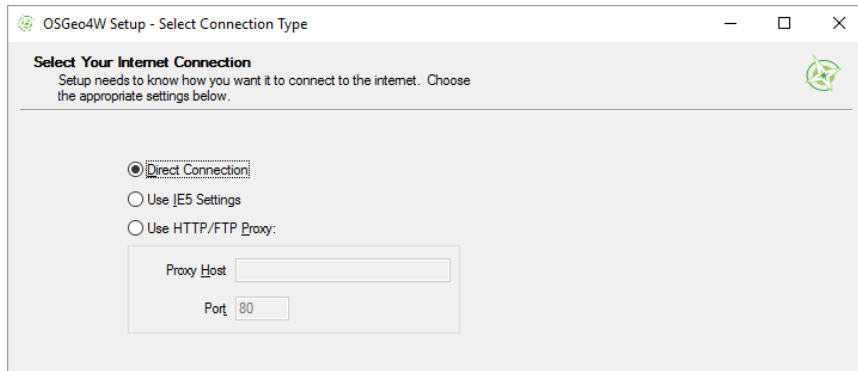


If needed navigate to the **Local Package Directory**, then click **Next**.

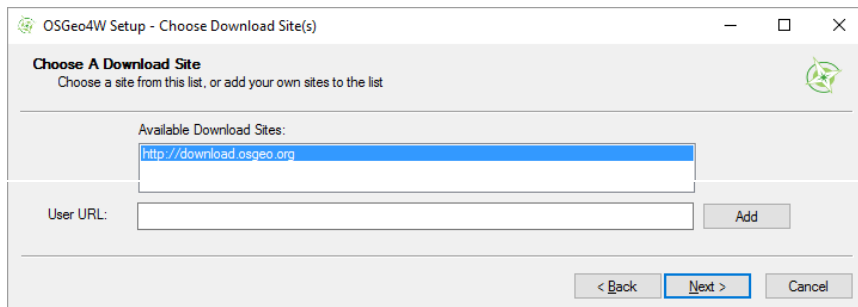


Select **Direct Connection**, click **Next**.

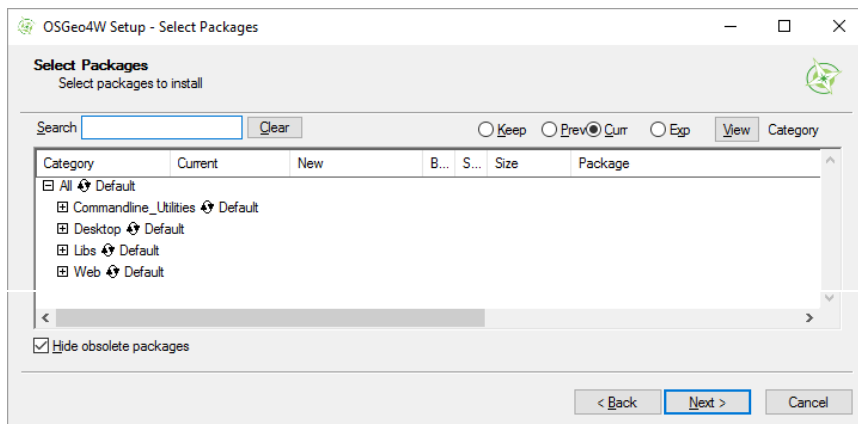
## Instructions for Open Source Object Oriented Classification



Click on the **available download site**, click **Next**.



Select Packages

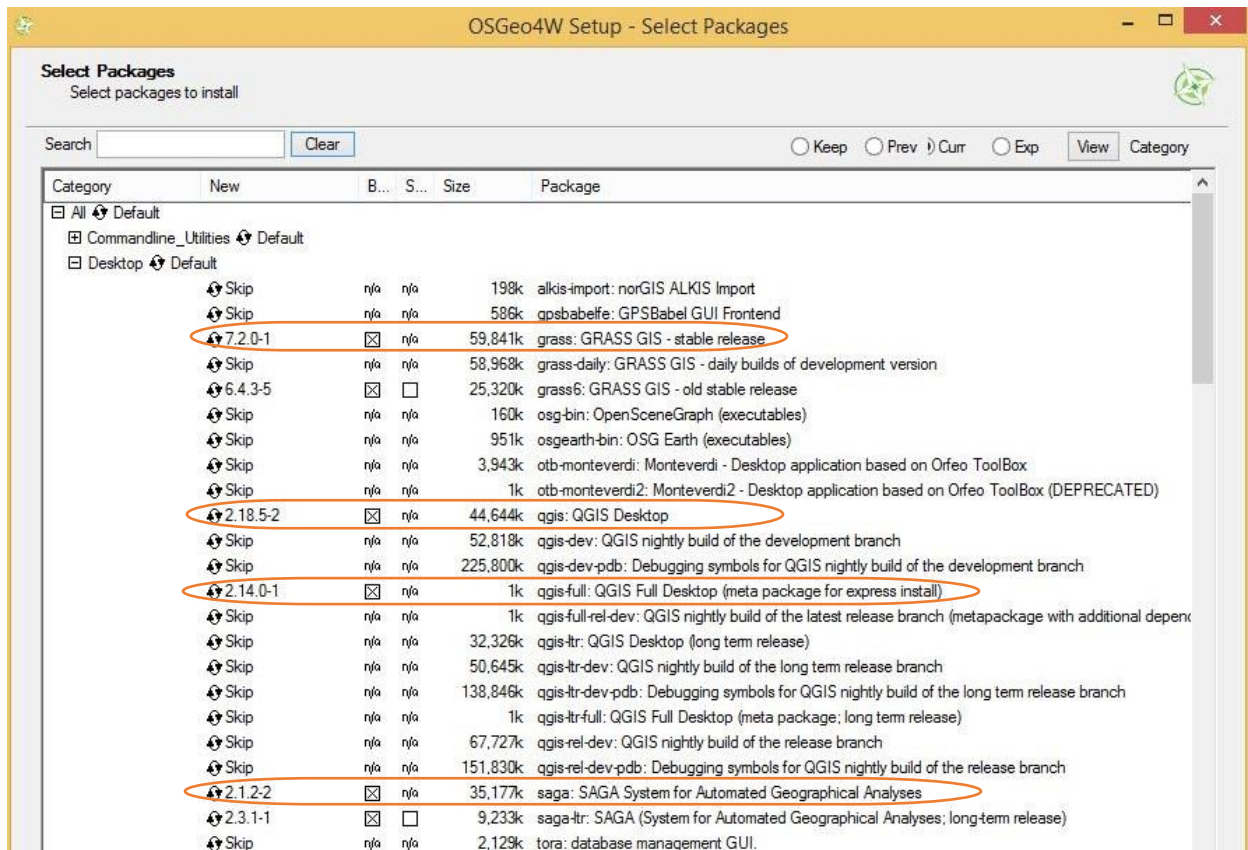


From the package selection page, expand the Desktop section and click just once on “Skip” for at least the following packages such that it shows the latest version number where it used to say “Skip”:

- grass: GRASS GIS – stable release
- qgis: QGIS Desktop
- qgis-full: QGIS Full Desktop (meta package for express install)
- saga: SAGA System for Automated Geographical Analysis

Other packages with versions shown in the screenshot but not circled in red will automatically be added when you come to the later step where you select “Install these packages to meet dependencies.”

## Instructions for Open Source Object Oriented Classification

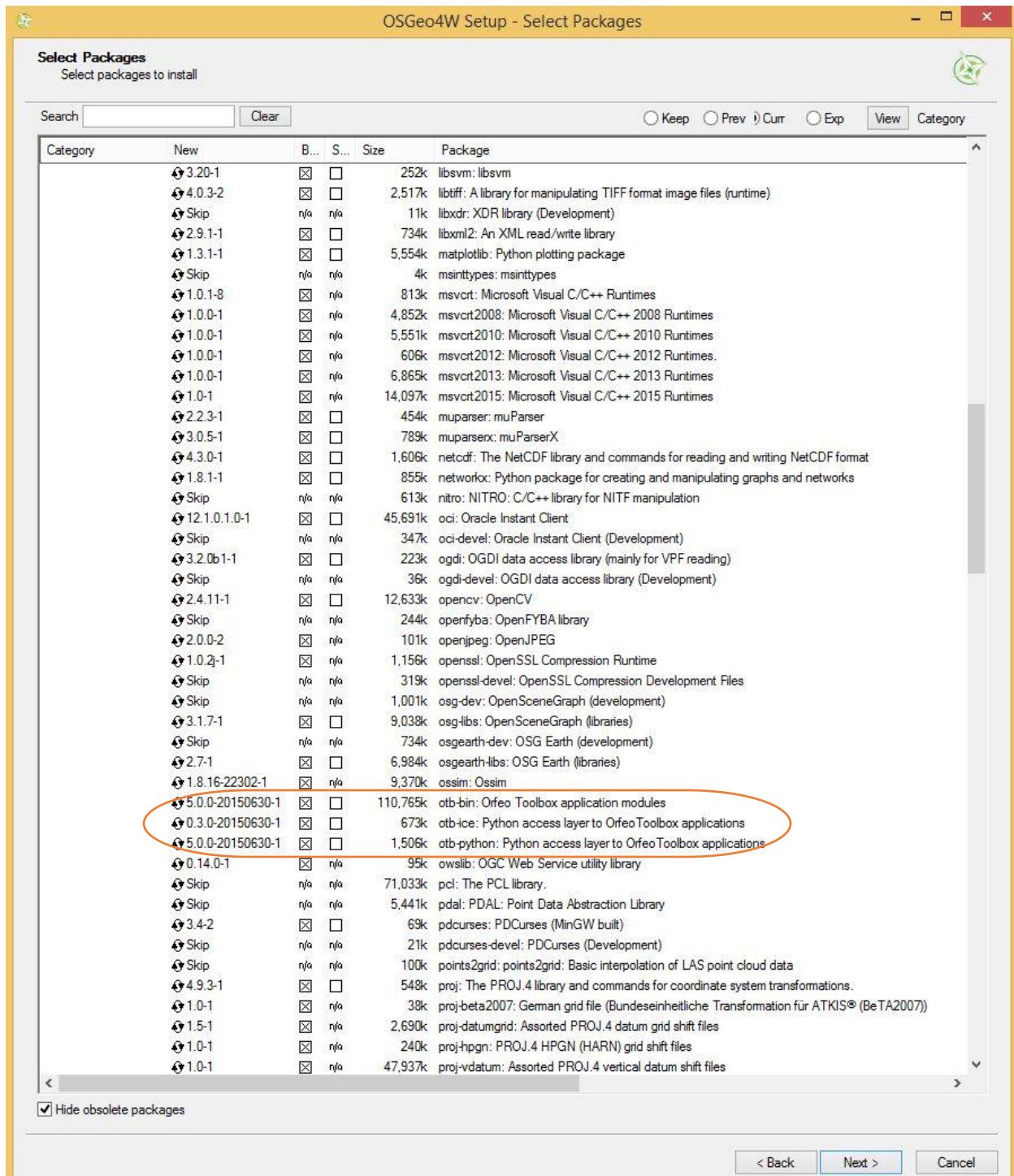


Then expand the Libs section to verify there is a check to install (Keep) Orfeo Tool Box (OTB) at least:

- otb-bin: Orfeo Toolbox application modules
- otb-ice: Python access layer to Orfeo Toolbox applications
- otb-python: Python access layer to Orfeo Toolbox applications
- pyspatialite: Python interface to SQLite 3 + Spatialite
- python-numpy: Num Py: Array processing for numbers, strings, records and objects
- python-rpy2: python bindings to R 3.0.2
- python-scipy: Scientific library for Python
- python-shapely: Geospatial geometries, predicates, and operations for Python.

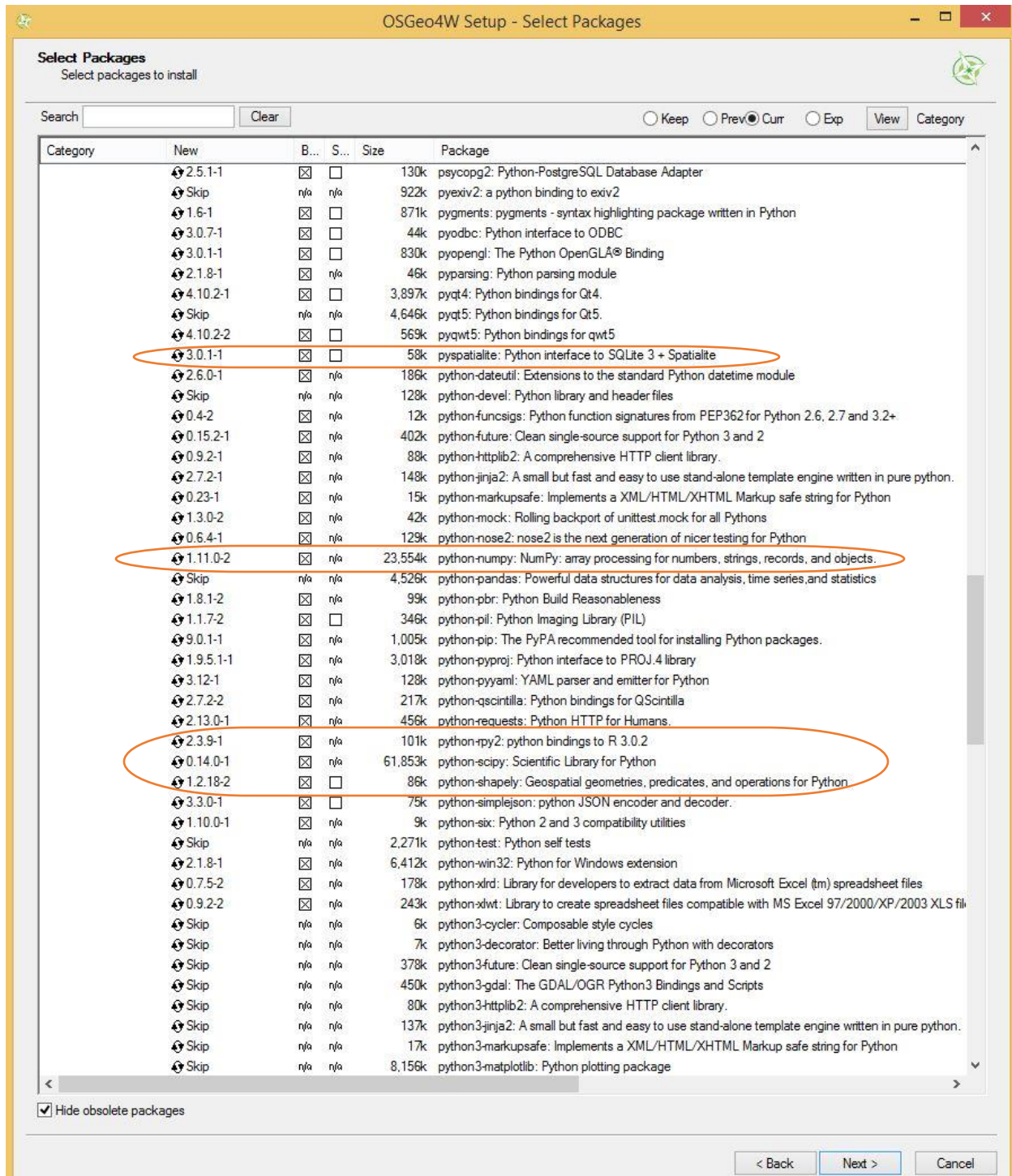
If space is not a limiting factor, you may consider a complete installation. The Install All option will also take longer time to download and install. Click on the above screen under “**Category**” immediately to the right of “**All**” the revolving arrows, and cycle through the iterations until the “**Install**” option is next to All. This will install all the options guaranteeing that you have ORFEO Tool Box and all the other tools available at the time of the release, you can skip to the end of this chapter if you choose to go with a complete installation, otherwise continue through the options.

## Instructions for Open Source Object Oriented Classification





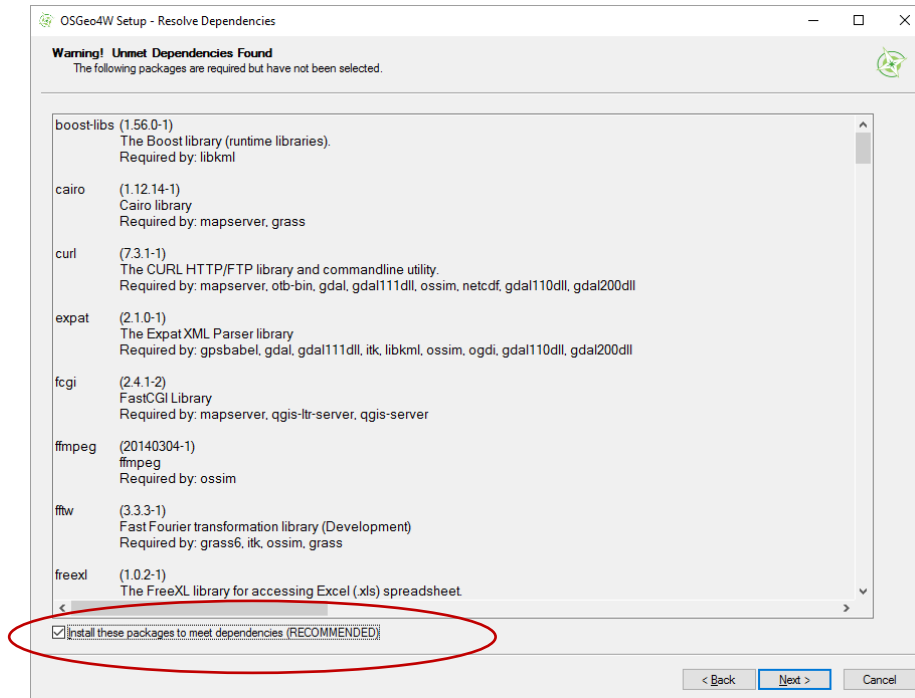
## Instructions for Open Source Object Oriented Classification



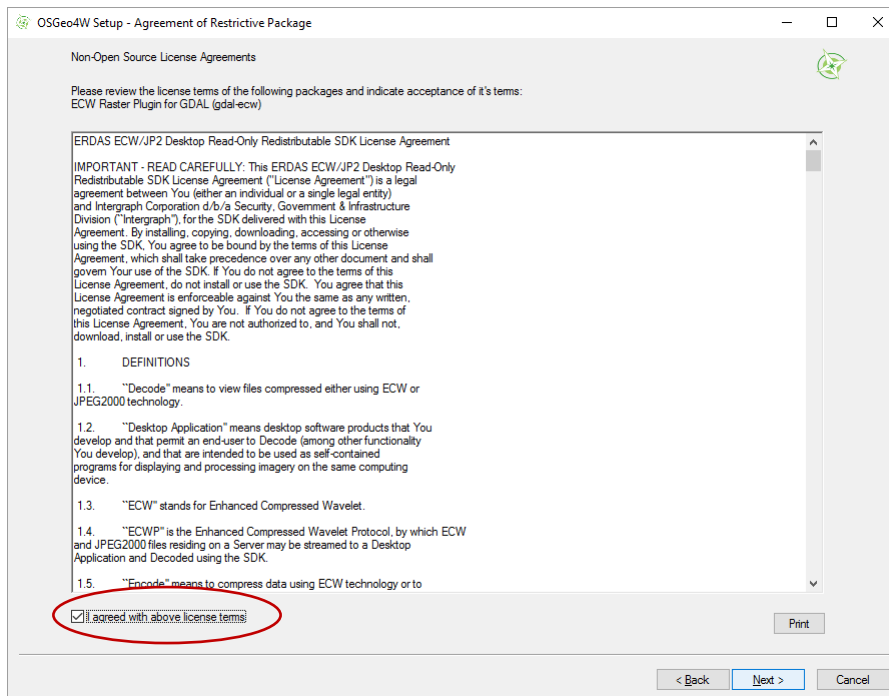
Click **Next**.

On the new screen, make sure that the check mark is on for: Install these packages to meet dependencies.

## Instructions for Open Source Object Oriented Classification

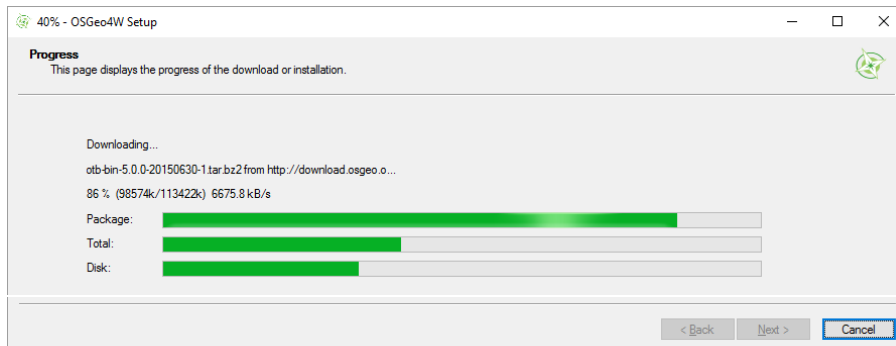


Click **Next** and agree with the license terms for the software and plugin options selected.



The software will start downloading all the options and installing them.

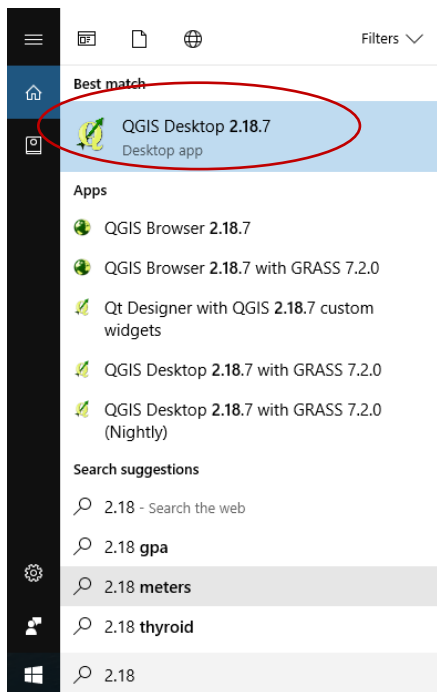
## Instructions for Open Source Object Oriented Classification



After the successful installation, click **Finish**.

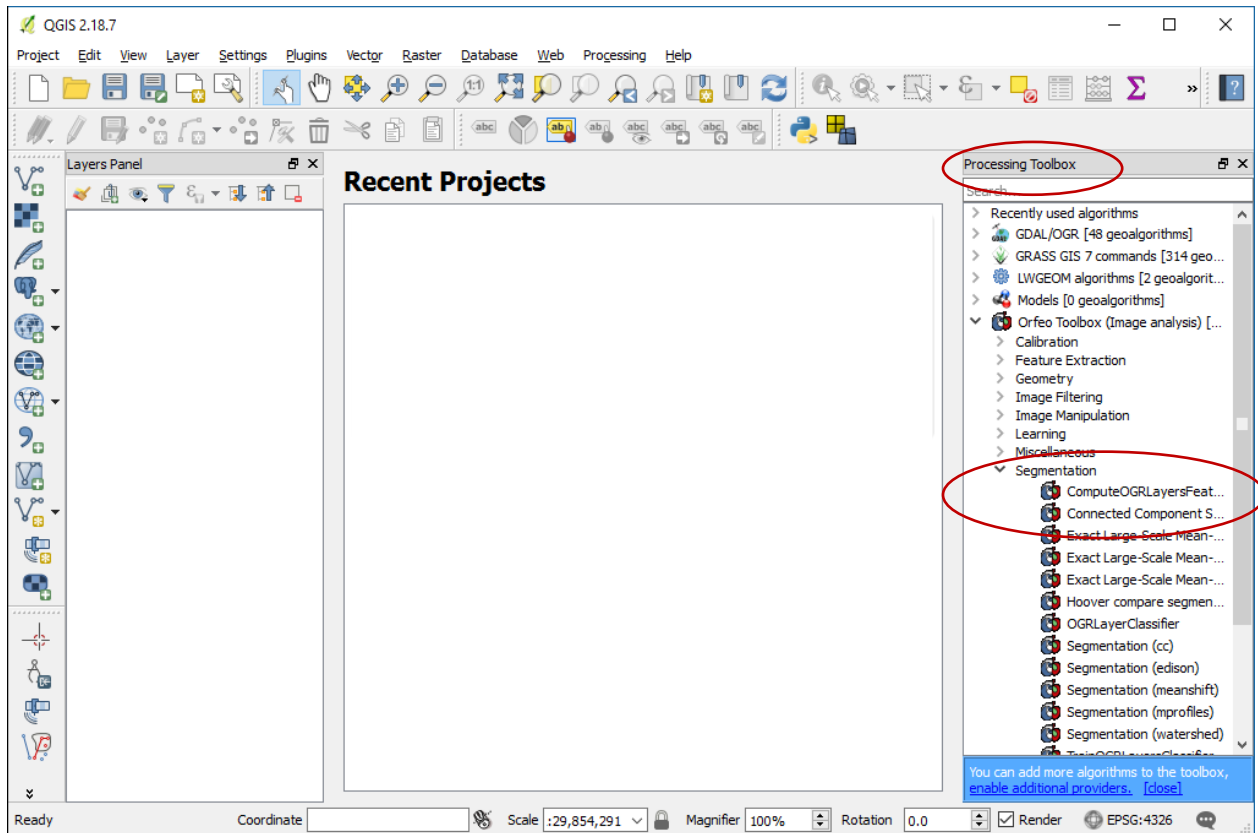
### 1.3 Verify the installation of your QGIS Software

In your platform applications, **open** the freshly installed version of QGIS desktop.



Open the **Processing toolbox**, under **Orfeo Toolbox > Segmentation**, you will find several segmentation tools that can be used for Object Oriented Classification of Remotely Sensed Images.

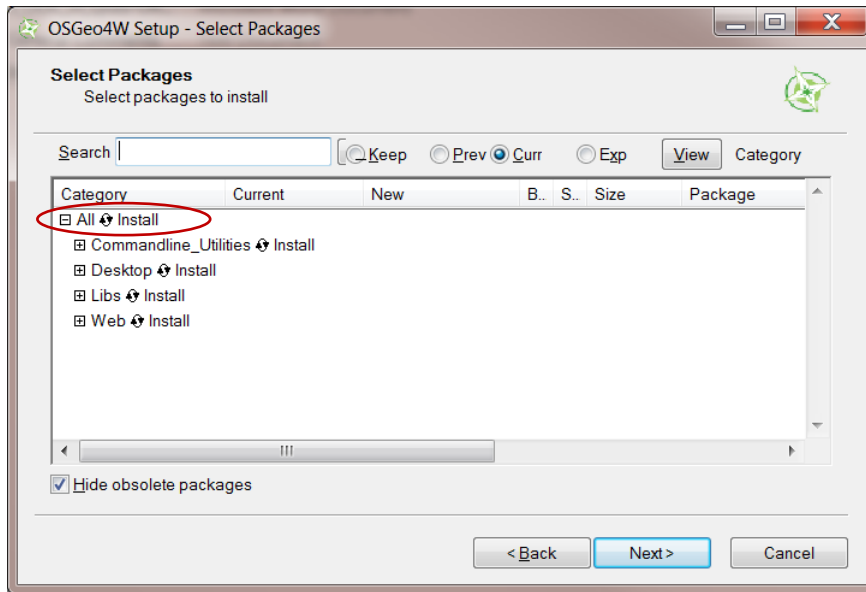
## Instructions for Open Source Object Oriented Classification



If the **Orfeo Toolbox > Segmentation** is not there, or if it is empty, you may need to re-run the **OSGEO4W Network Installer (64 bit)** to revise your options. As versions change, these options change as well; if in doubt, one sure way to get the Orfeo Toolbox is to do a complete install.

Go to your downloads and run **OSGEO4W Network Installer (64 bit)** again. Near the end of the installation, at the step where you need to Select Packages, pay attention to the options. Under the Category column, click next, immediately to the right of "**All**" and cycle through the iterations until the "**Install**" option is next to **All**, as below.

## Instructions for Open Source Object Oriented Classification



Installing all the packages in a complete installation will take more space in your system, but also will ensure that you have all the tools, including the segmentation and classification tools needed to perform an Object-Oriented Classification with OTB in QGIS.

## 2 Earth Explorer Select and Download an image to use in your Object Oriented Classification

Earth Explorer is a user interface to query and order satellite images and other U.S. Geological Survey products. The user can specify a search criteria, select datasets, and browse the results before downloading the image(s) and product(s) needed for their project. This section shows how to identify and download remotely-sensed imagery freely available to use in your Object-Oriented Classification.

### 2.1 Landsat 8.

Landsat 8 is the latest NASA satellite in a series that has produced an uninterrupted multispectral record of the Earth's land surface since early 1970s. Landsat 8's return (re-visit) period is 16 days. There are two instruments on board, the Operational Land Imager (OLI, 8 bands at 30 m spatial resolution plus one panchromatic band at 15 m) and the Thermal Infrared Sensor (TIRS, 2 bands at 100 m, re-sampled and co-registered at 30 m). Data collected by these sensors is available for download at no charge via EarthExplorer. Additionally, there are other interfaces, e.g. GloVis and LandsatLook Viewer. Landsat data is typically available within 24 hours of reception. Questions about Landsat can be answered in the USGS' FAQ list (<https://Landsat.usgs.gov/frequently-asked>), the Landsat 8 handbook (<https://Landsat.usgs.gov/Landsat-8-l8-data-users-handbook>), via e-mail ([custserv@usgs.gov](mailto:custserv@usgs.gov)) or by phone (800-252-4574).

For this example, we will download a Landsat 8 image in the Desert Landscape Conservation Cooperative (DLCC) near Cd. Juarez - El Paso on the US/Mexico border. Calibrated imagery is available through Earth Explorer.

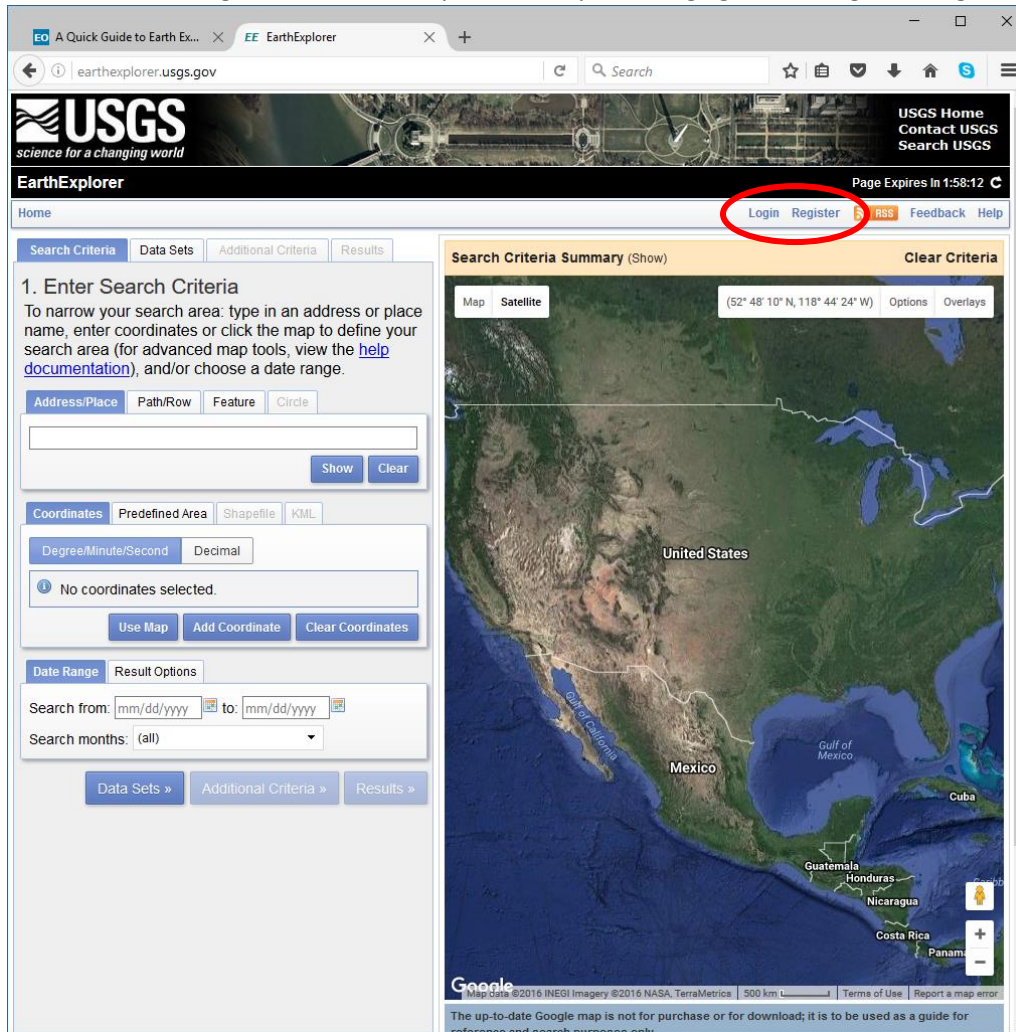
### 2.2 Create your login at EarthExplorer

The USGS' Earth Explorer website allows the user to custom tailor search parameters for Landsat data. After entering the website, <http://earthexplorer.usgs.gov> (new users must create a login, access to the site and the data is still free), you can either search for data by selecting a region on a map, by entering coordinates, or by entering a place name. The user can also select specific parameters such as acceptable cloud cover percentages and the range of desired acquisition dates.



## Instructions for Open Source Object Oriented Classification

In a browser, navigate to the site <http://earthexplorer.usgs.gov>, and log in or register a new account.



### 2.3 Search and Select the data you need

On the left panel, the first tab is the Search Criteria which contains three sets of search options.

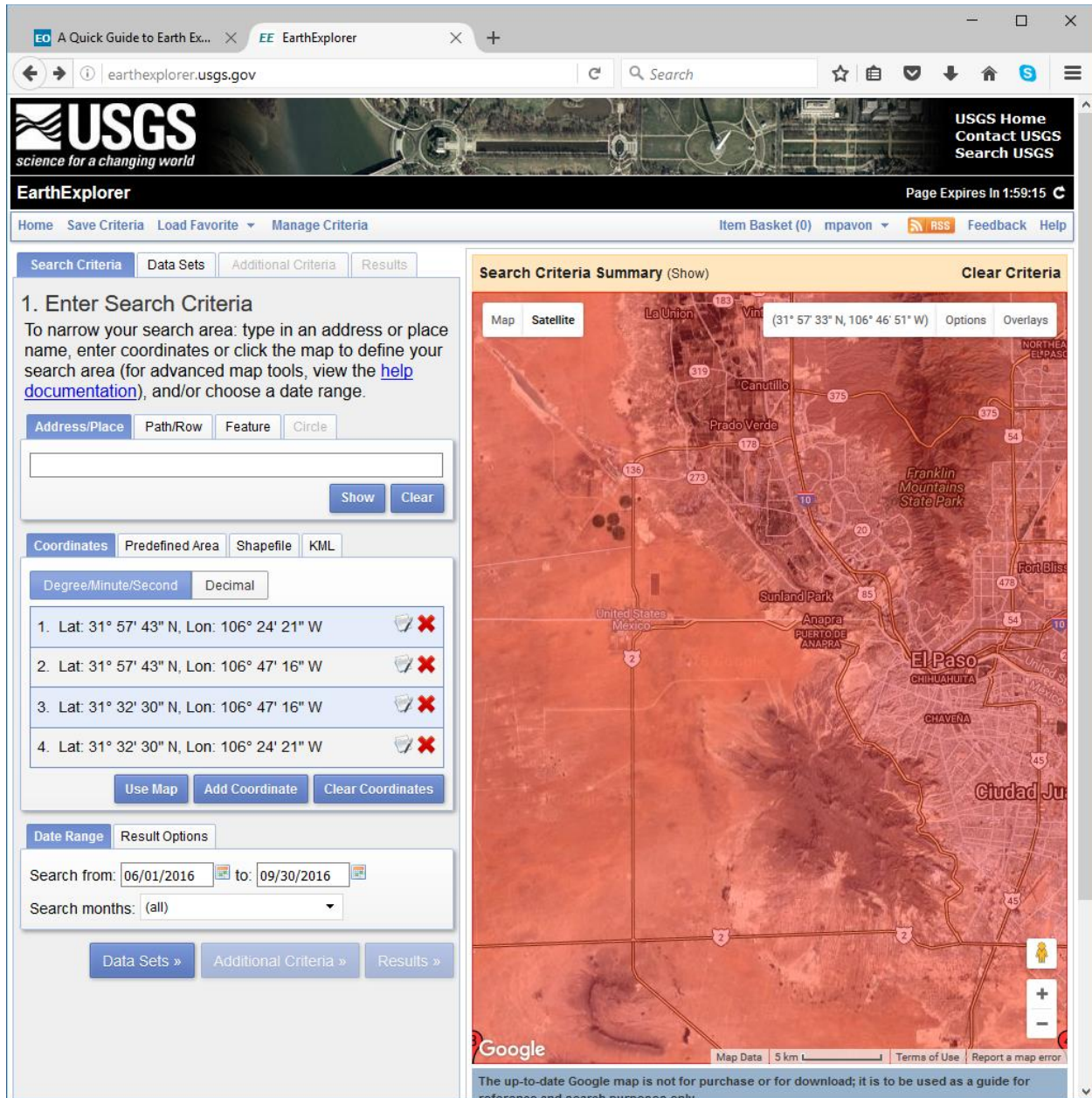
The first set of options allows the user to identify a particular address or place name, a path-row combination or the name of a feature in your Area of Interest (AOI) you can define it here.

The second set of options permits to type coordinates or to use a shapefile or KML with a pre-defined area as input. Additionally, the user can use the “Map” option by just clicking on the map and populate a set of coordinates.

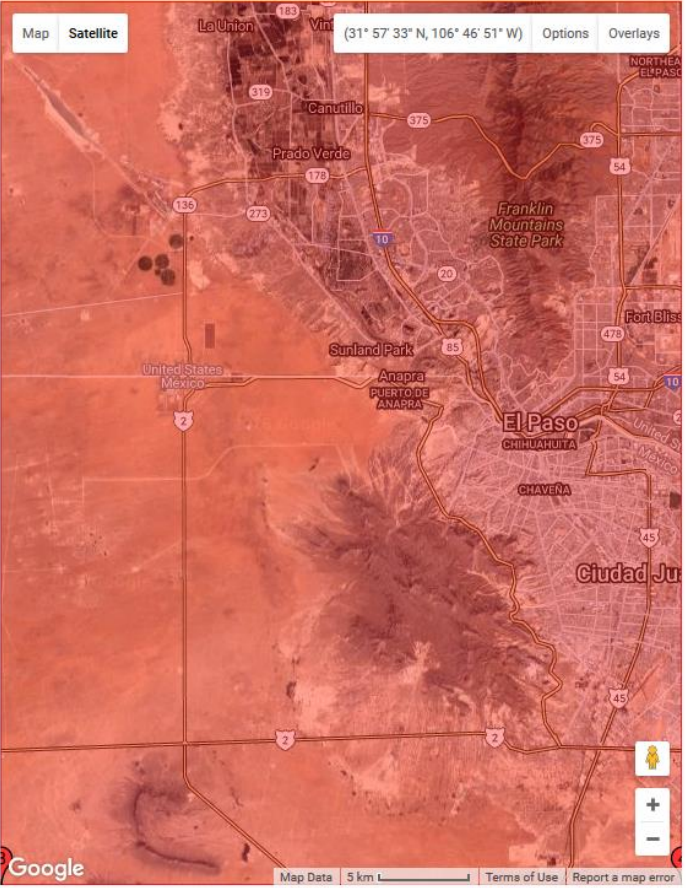








The third set of options during search criteria allows for date ranges. If the user has a particular year or month in which field data collection took place, a year that was not too dry or too wet, or other dates, in particular, a range can be specified here.

## Instructions for Open Source Object Oriented Classification

Zoom into an area west of Cd. Juarez - El Paso, on the international border and click “Use Map” that will take the coordinates of your view as part of the search criteria. For dates, use the summer of 2016, from 06/01/2016 to 09/30/2016 for this tutorial. Notice that in the Login-Register section now a user is defined.



The screenshot shows the USGS Earth Explorer website. The top navigation bar includes the USGS logo, a search bar, and links for Home, Save Criteria, Load Favorite, and Manage Criteria. The main content area is divided into two sections: "Search Criteria" and "Data Sets". The "Search Criteria" section is active and displays a "Search Criteria Summary" table with four entries. The first entry is selected, showing coordinates (31° 57' 43" N, 106° 24' 21" W) and a date range from 06/01/2016 to 09/30/2016. The "Data Sets" section is partially visible on the right, showing a list of datasets including "Aerial Imagery", "Lidar Elevation", and "AVHRR Phenology".

Search Criteria Summary (Show)		Clear Criteria	
Map	Satellite	(31° 57' 33" N, 106° 46' 51" W)	Options Overlays
			
1. Lat: 31° 57' 43" N, Lon: 106° 24' 21" W  			
2. Lat: 31° 57' 43" N, Lon: 106° 47' 16" W  			
3. Lat: 31° 32' 30" N, Lon: 106° 47' 16" W  			
4. Lat: 31° 32' 30" N, Lon: 106° 24' 21" W  			
<input type="button" value="Use Map"/> <input type="button" value="Add Coordinate"/> <input type="button" value="Clear Coordinates"/>			
<input type="button" value="Date Range"/> <input type="button" value="Result Options"/>			
Search from: 06/01/2016 to: 09/30/2016			
Search months: (all)			
<input type="button" value="Data Sets »"/> <input type="button" value="Additional Criteria »"/> <input type="button" value="Results »"/>			

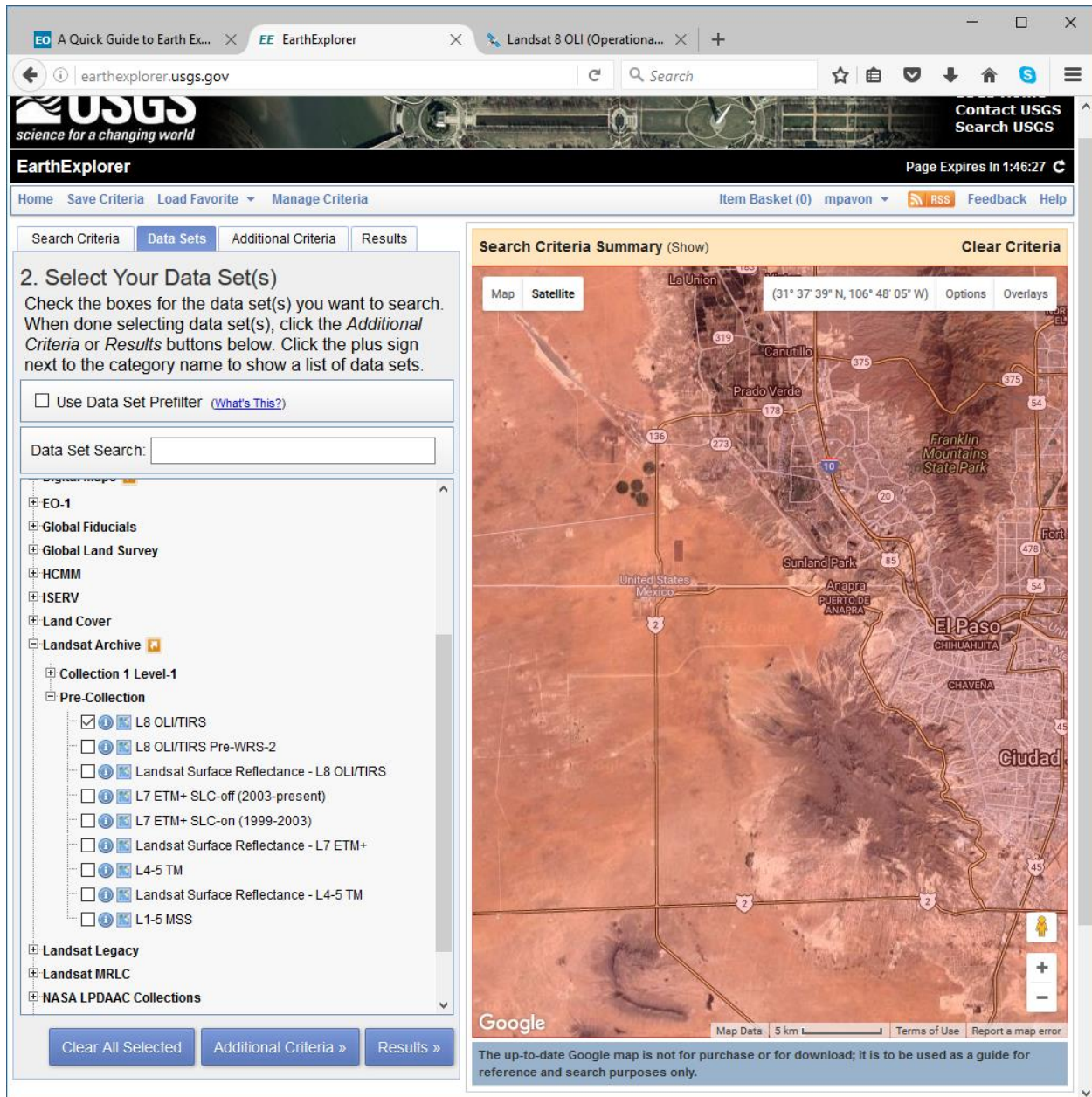
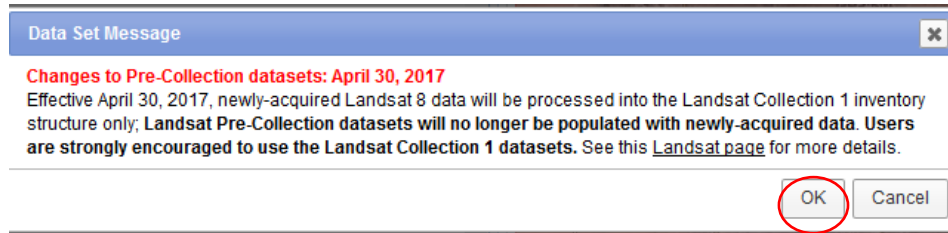
The second tab is the “Data Sets” page. Many datasets are available, from “Aerial Imagery” to “Lidar Elevation” to “AVHRR Phenology” that may be useful for DLCC. In this case, we will look for a Landsat 8 image.

Go to the second tab, “Data Sets”. Expand “Landsat Archive”, Under “Pre-Collection” put a checkmark on “L8 OLI/TIRS”. Those are the imagery sensors for Landsat 8: OLI (Operational Land Imager) and TIRS (Thermal Infrared Sensor). For this tutorial, we will work only with the OLI data. You may get a warning



## Instructions for Open Source Object Oriented Classification

for newer Landsat datasets after April 30, 2017 redirecting you to “Collection 1 datasets”, acknowledge the message by clicking **OK** and continue.



Click the Additional Criteria Tab. Scroll down to “Cloud Cover” and pick “Less than 10%”. That should limit your results to those that have minimal cloud cover. Click on Results Tab. The user can get a list of

## Instructions for Open Source Object Oriented Classification

available images that fit the criteria with thumbnails and links to more information. The thumbnails aren't big enough to show much, so click on one to see a slightly larger image.

The screenshot shows the EarthExplorer website interface. The top navigation bar includes links for Home, Save Criteria, Load Favorite, and Manage Criteria. The main content area is divided into two sections: Search Criteria Summary and Search Results.

**Search Criteria Summary (Show) Clear Criteria**

Map Satellite (31° 52' 35" N, 106° 41' 07" W) Options Overlays

**4. Search Results**  
If you selected more than one data set to search, use the dropdown to see the search results for each specific data set.

Show Result Controls

**Data Set** Click here to export your results »

L8 OLI/TIRS

« First « Previous 1 Next » Last »

Displaying 1 - 5 of 5

1 Entity ID: LC80330382016262LGN00  
Coordinates: 31.74231, -106.72691  
Acquisition Date: 18-SEP-16  
Path: 33  
Row: 38

2 Entity ID: LC80330382016246LGN00  
Coordinates: 31.74206, -106.72851  
Acquisition Date: 02-SEP-16  
Path: 33  
Row: 38

3 Entity ID: LC80330382016230LGN00  
Coordinates: 31.74216, -106.71758  
Acquisition Date: 17-AUG-16  
Path: 33  
Row: 38

Entity ID: LC80330382016198LGN00  
Coordinates: 31.74217, -106.72923  
Acquisition Date: 16-JUL-16

View Item Basket » Submit Standing Request »

The map view shows a satellite image of the El Paso area, New Mexico, with labels for various locations including Canutillo, Prado Verde, Sunland Park, Anapra, Puerto de Anapra, El Paso, Chihuahua, and Ciudad. The map includes a scale bar (5 km) and a disclaimer: "The up-to-date Google map is not for purchase or for download; it is to be used as a guide for reference and search purposes only."

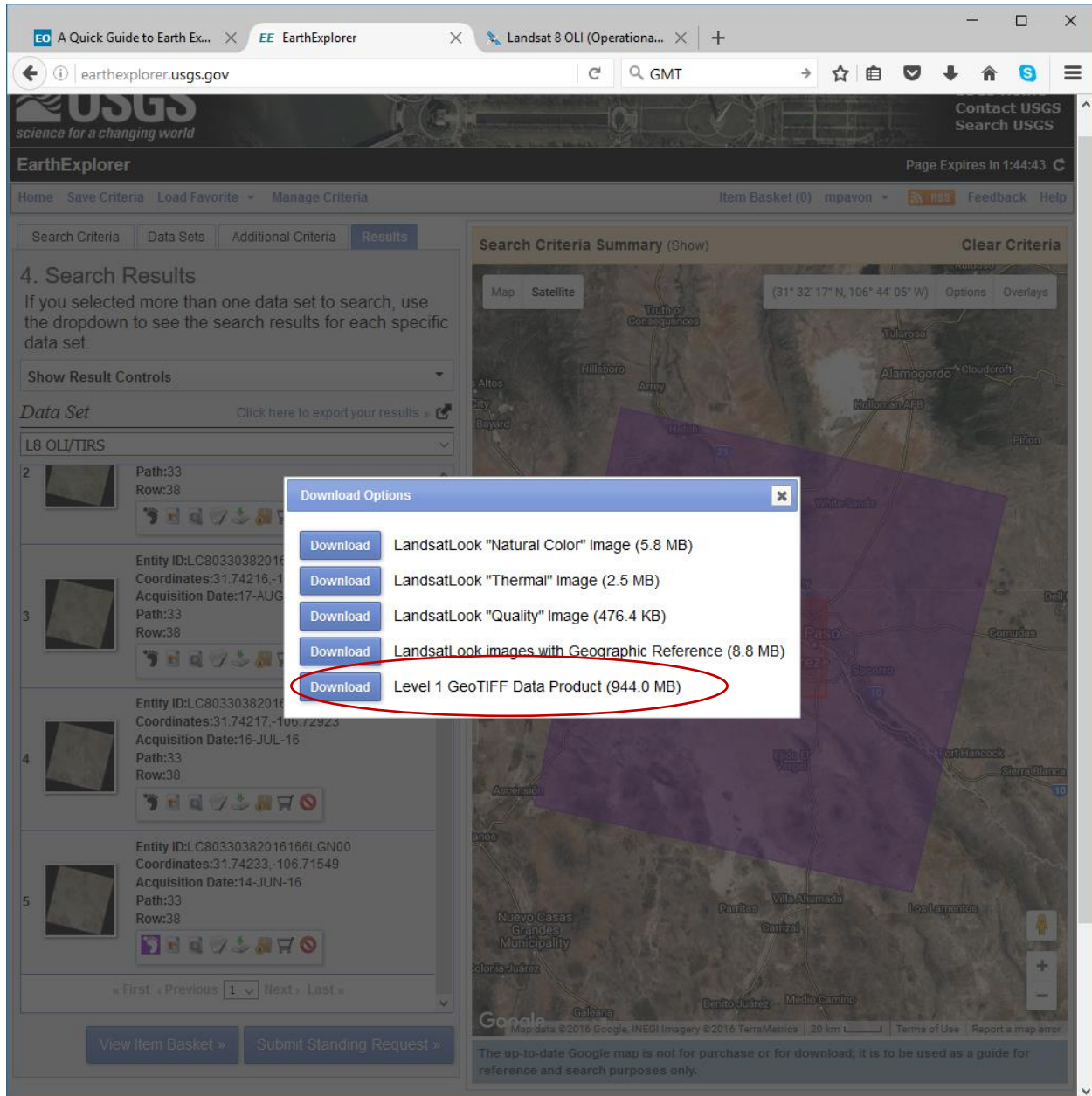
When clicking on one image a new window opens allowing the user a better view of the image and some metadata. Go through all of them until you get familiar with the results. Looks like the granule for 06/14/2016 has minimal cloud cover and was taken about 11 AM local time (El Paso is GMT -6).

### 2.4 Download your data



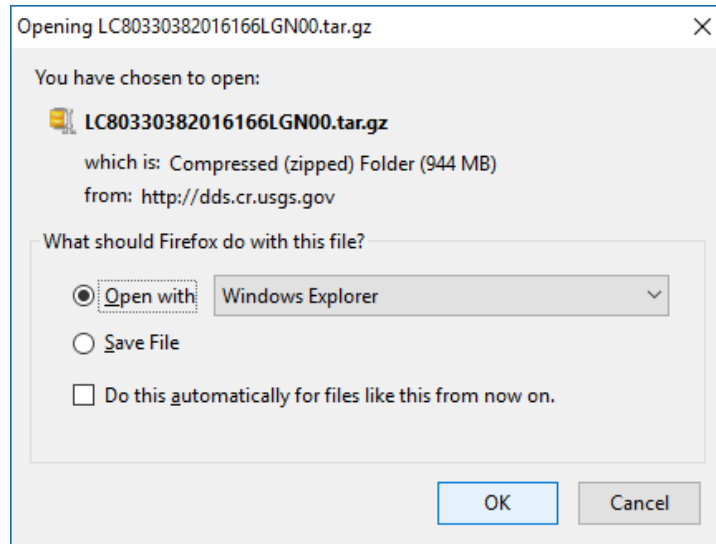
## Instructions for Open Source Object Oriented Classification

Hovering over the different icons for each image shows a little text of what they do; in order they are: show footprint, show browse overlay, compare browse, show metadata and browse, download options, add to bulk download, order scene and exclude scene from results. Explore the icons by clicking on them. Click on download options for Entity ID:LC80330382016166LGN00.

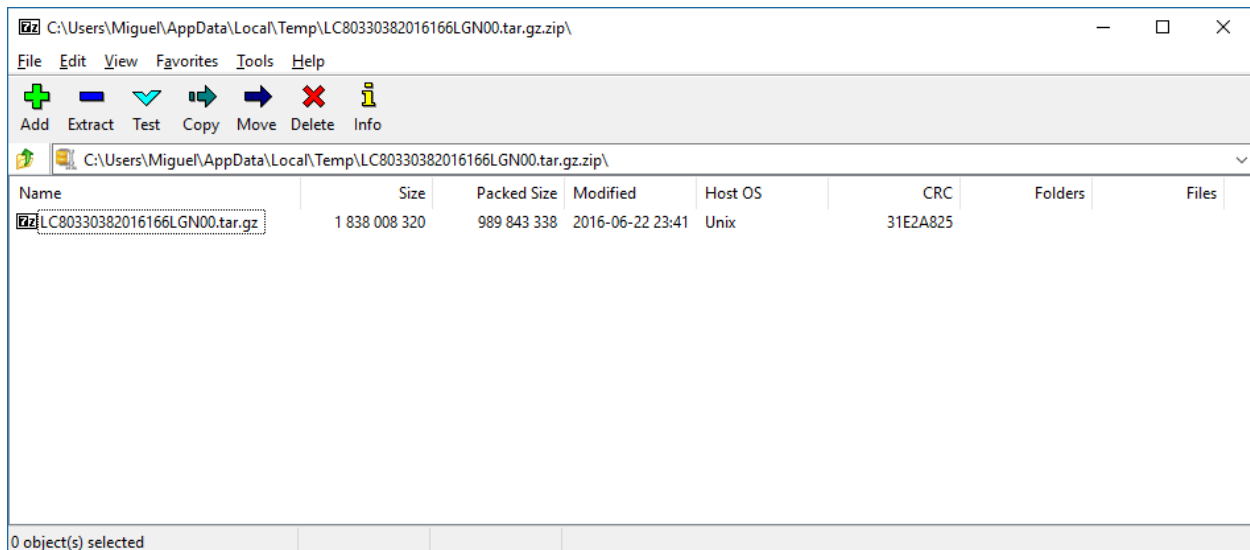


The bigger option, "Level 1 GeoTIFF Data Product" is the one that provides a full-resolution, uncompressed product. The Level 1 Product is terrain-corrected, geolocated, calibrated data—a bundle of 16 bit, single-channel GeoTIFFs. The other ones are just full resolution JPEG previews to check on the image quality. Click on the "Download" button next to it. A compressed file will start downloading.

## Instructions for Open Source Object Oriented Classification

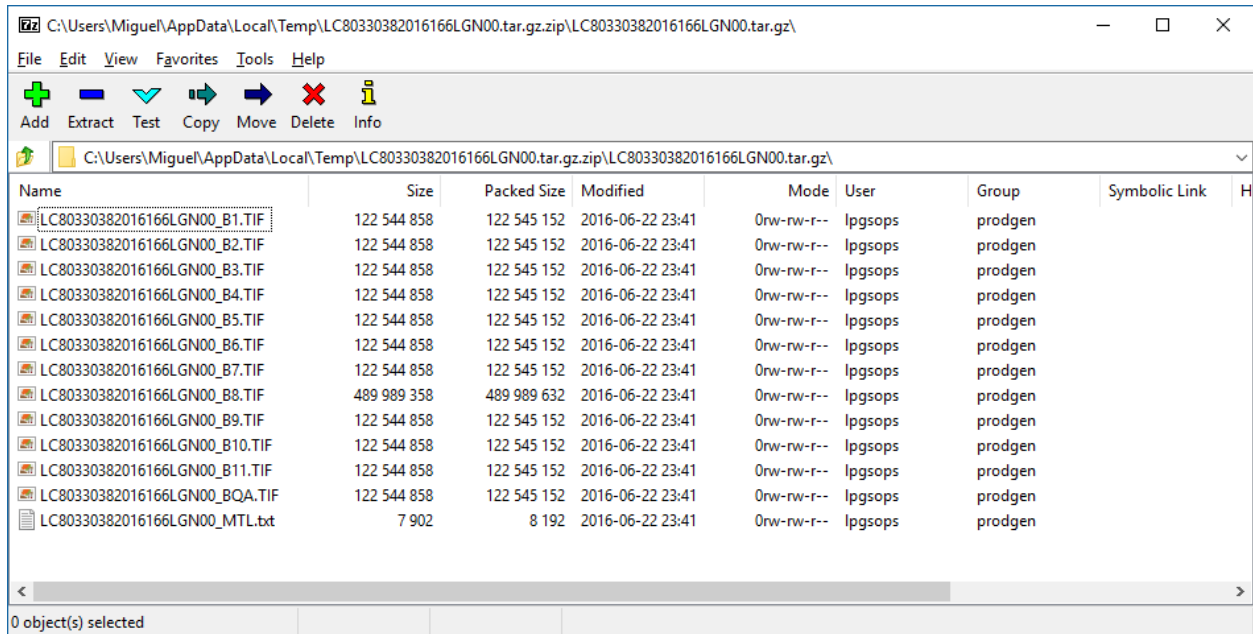


Choose a program with which you can open a compressed file (WinZip (<http://www.winzip.com>), 7Zip (<http://www.7-zip.org>), etc.) Depending on your internet connection, this may take from one to several minutes. We used 7-Zip as a way to open it (7-Zip is open source software available at <http://www.7-zip.org/>, and there are apps and utilities also available for Mac users: command line, RAR extractor (<http://www.unrarx.com>), Unarchiver (<http://unarchiver.c3.cx/unarchiver>), etc.). Tar.gz is a compressed file inside of a compressed file.

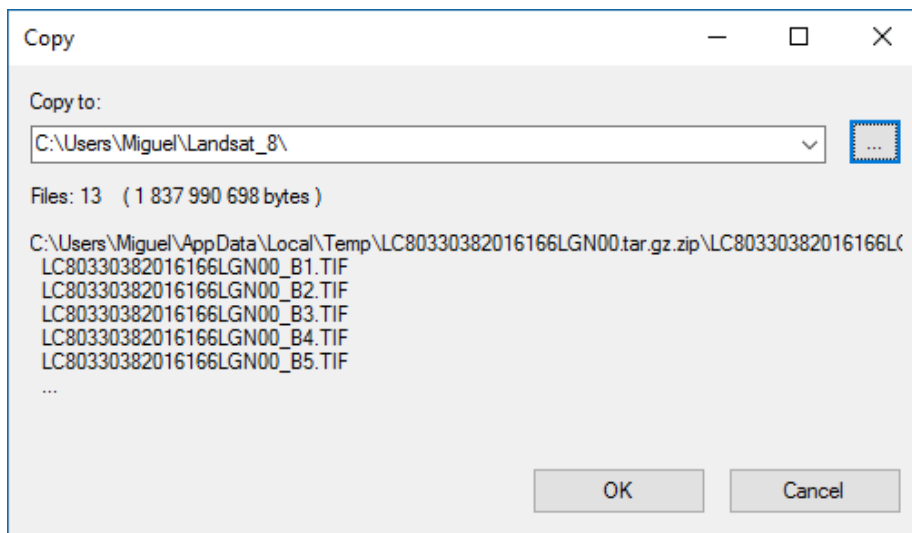


Double click on the gz file to uncompress the gz file into its components.

## Instructions for Open Source Object Oriented Classification



Click the extract button and download the individual bands 1-11, quality assessment image (BQA) and metadata (txt) into a local folder.



Now the Landsat 8 data is on your local hard drive. In the next tutorial we will work on making it into an image stack and use ground samples to perform Object Oriented Classification on it.

### 3 Object Oriented Classification with QGIS with ORFEO Toolbox.

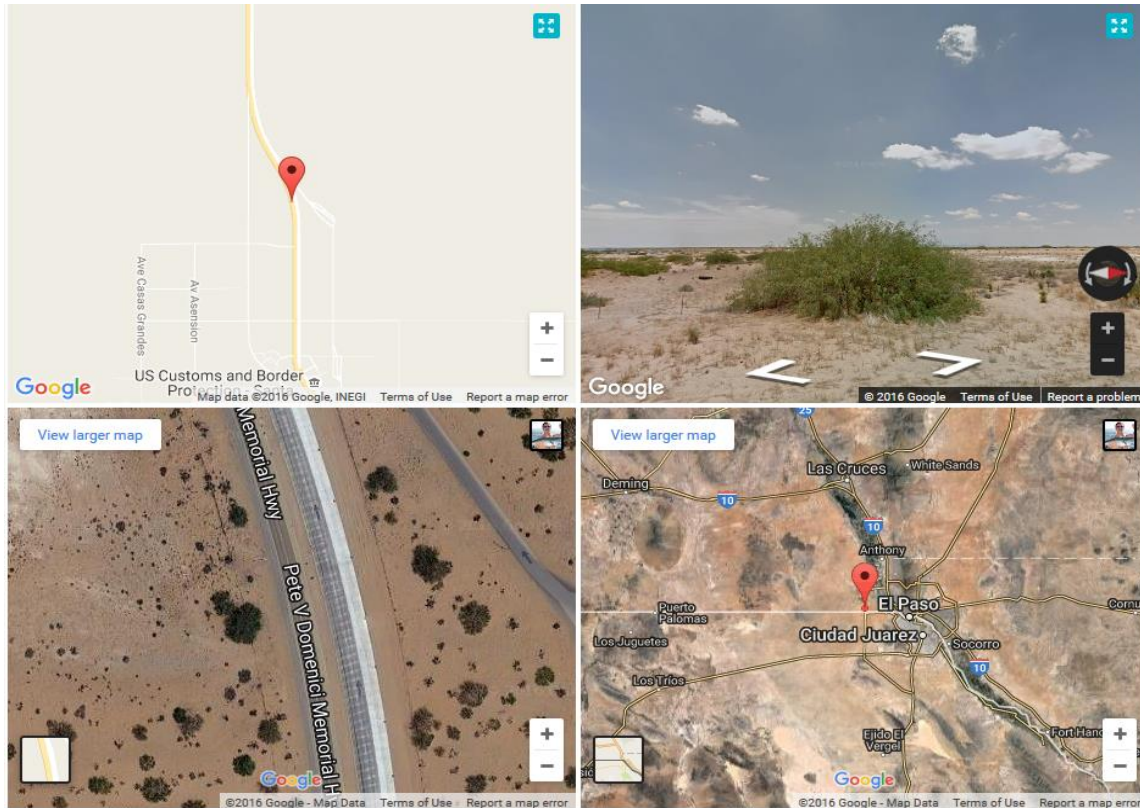
The Orfeo Toolbox (or OTB) is a library for remote sensing image processing, Object Oriented Classification (OOC) is included in it. The OTB can be described as a plugin for QGIS to perform image analysis.

#### 3.1 Sample polygons and ground-truthing

For OOC we need to have training polygons that define the land cover classes that we hope to create, which are called the objective classes. You create each training polygon around a known land cover class to “teach” the classification algorithm what that land cover looks like. A training polygon or point is a reference to a place on the ground where you are positive of the land cover at the time of collection, recording attributes of each check point should be the land cover, time/date of collection and X,Y coordinates (in a defined coordinate system). Additionally, we need to create the image stack of all the Landsat 8 bands. An image stack is the combination of 11 different ranges of frequency of light available for one image, e.g. blue, green, red, near infrared, etc., layered on top of one another on the map to create a multilayer image) and the ground truthing check-points to assess the accuracy of the classification.

Ideally both the training polygons and the check points are field-verified and documented with photographs. Field collection of ground truthing points (or polygons) should also be temporally close to the date of collection of the remotely-sensed imagery (we will be using Landsat 8 in this exercise). This information gives us the ground truthing for what was actually on the ground at the time of image collection. Training polygons and check points should be distributed throughout the extent of the image stack and should include a recommended minimum of 10 polygons and points per land cover class, to give enough statistical diversity to the classifier. In real projects, field work is needed to capture ground truthing information, i.e. determining what the correct land use class is based on visiting the polygon’s location. In our case, as field photographs were not collected or readily available, we used as an instructional tool imagery from google maps street view, high resolution aerial imagery and user photographs. An example of such photographs and locations near the Texas/Mexico border is seen below:

## Instructions for Open Source Object Oriented Classification



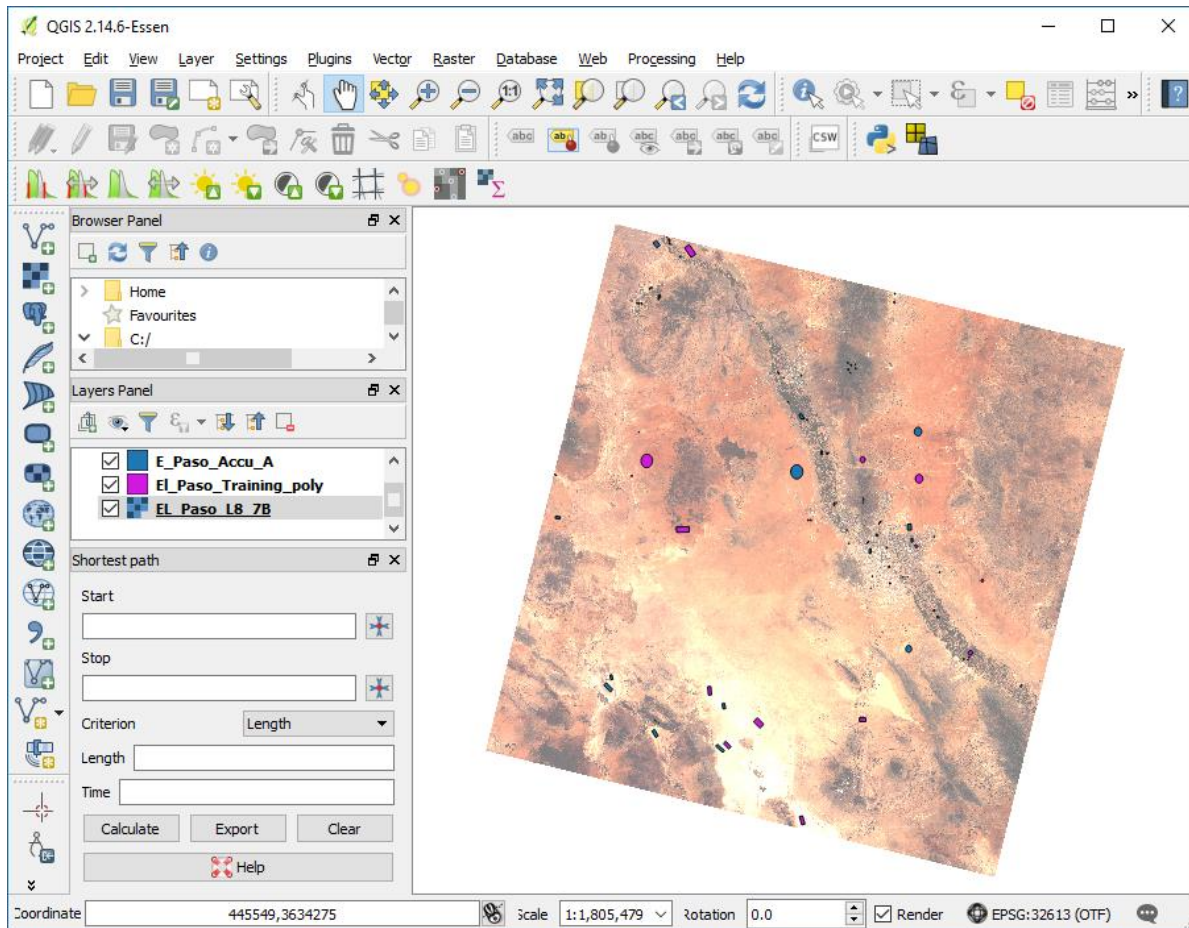
On the street view in Google maps, the date of the imagery is cited--it is best if the dating of the photograph is close to the remotely sensed image stack. This is not a condition that we can influence much, one more reason to include in your OOC plan to carry out field work ground truthing during the same time period as the date(s) of the LANDSAT image stack you plan to download.

For ground-truth polygon creation, once we have our El Paso Landsat 8, 7 bands, image stack, in QGIS, we need to add our ground truthing, collected with GPS, as points. We create a new polygon layer and we digitize polygons for each point based on the classes, location and extents on the field, documented in our notes and photographs. For this exercise, we used half of our points to create training polygons and the other half to create polygons that later will be used for the classification accuracy assessment as validation polygons.

For this exercise, close to 90 polygons were created for six classes: water, developed, grass and agriculture, brushland, barren land and forest. Training polygons will be used for training samples for use in the classification, while validation polygons will be used for accuracy assessment of the resulting classification. The polygons for the training samples (in a shapefile called "El\_Paso\_Training\_poly.shp") can be seen in magenta on top of the Landsat 8 image in natural color (Natural color combination, (4,3,2) refers to the Landsat bands displayed in red, green, blue (RGB) respectively). The polygons reserved for accuracy assessment are seen above the image in blue (reserved in a shapefile called: "El\_Paso\_accu\_A.shp").

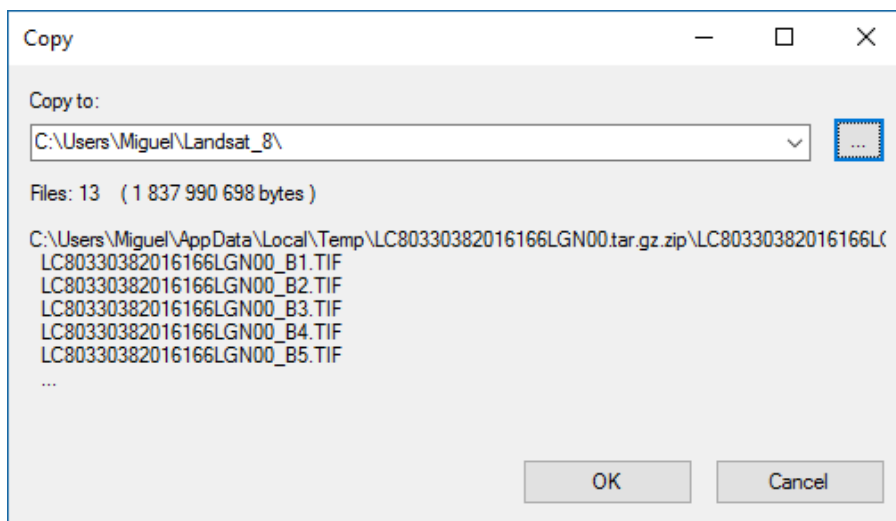


## Instructions for Open Source Object Oriented Classification



### 3.2 Create the image stack

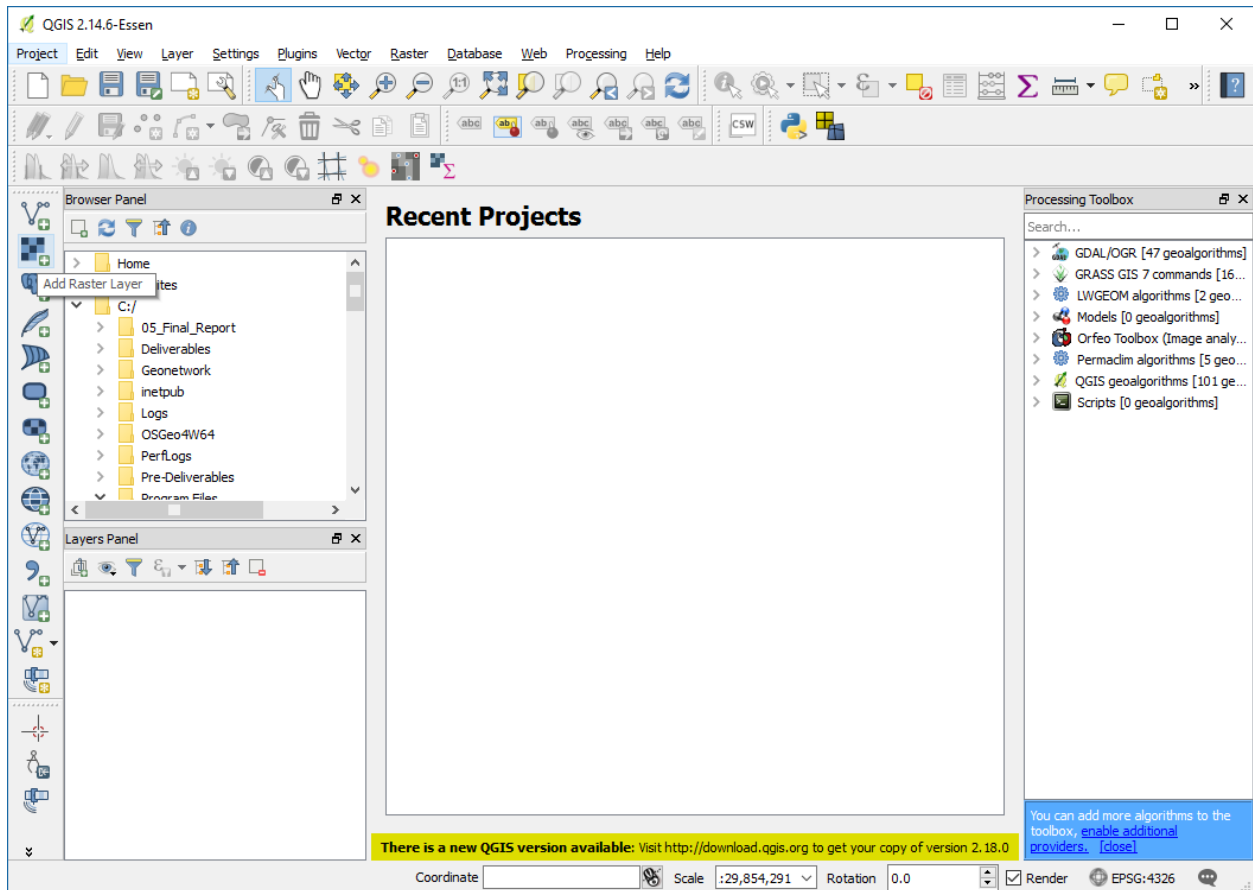
Go to the directory where you saved the individual bands of Landsat 8 imagery downloaded from Earth Explorer.





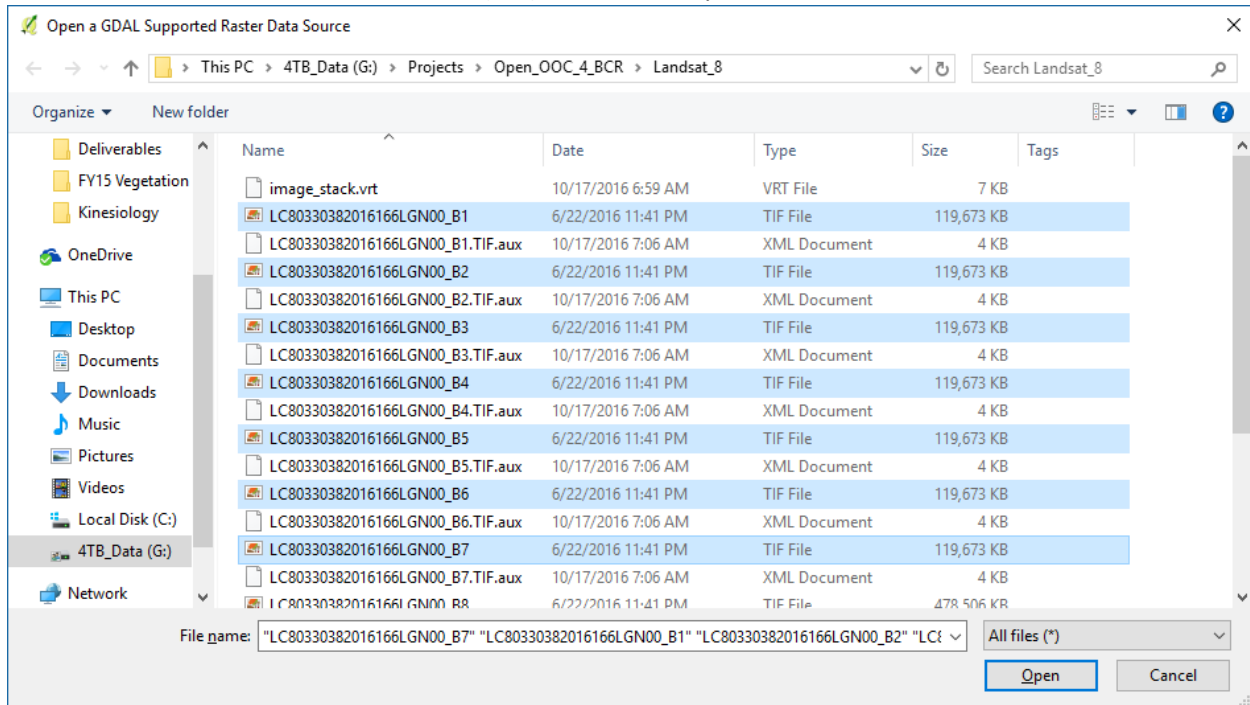
## Instructions for Open Source Object Oriented Classification

Open QGIS and add the layers with the **“Add Raster Layer”** button. Open the first 7 bands of the El Paso Landsat 8 image. They correspond to the first 7 bands of the OLI sensor without using the panchromatic.

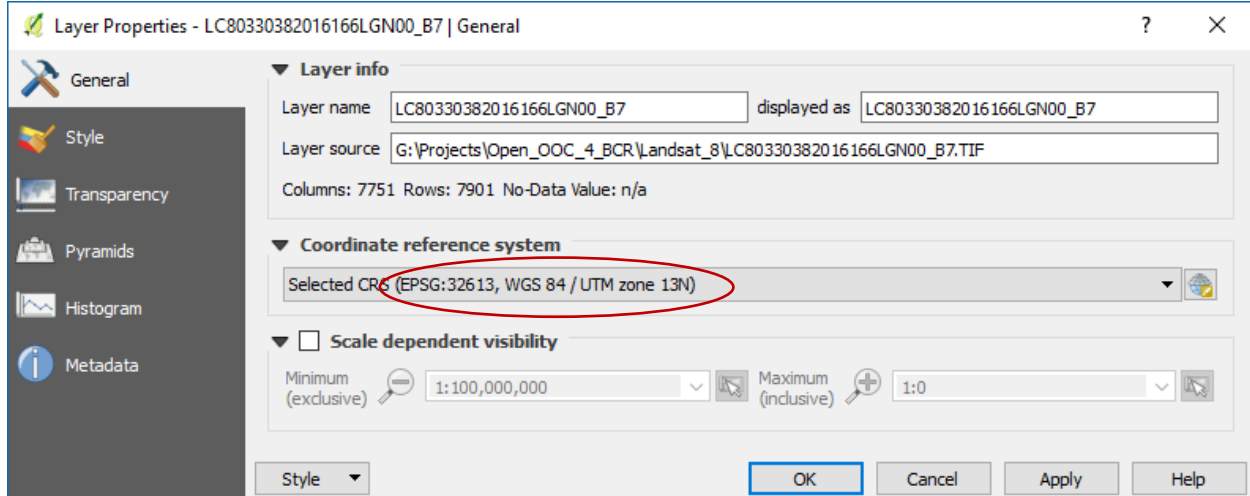


## Instructions for Open Source Object Oriented Classification

Select the 7 tif files for the first 7 bands (1-7), then click “Open”.

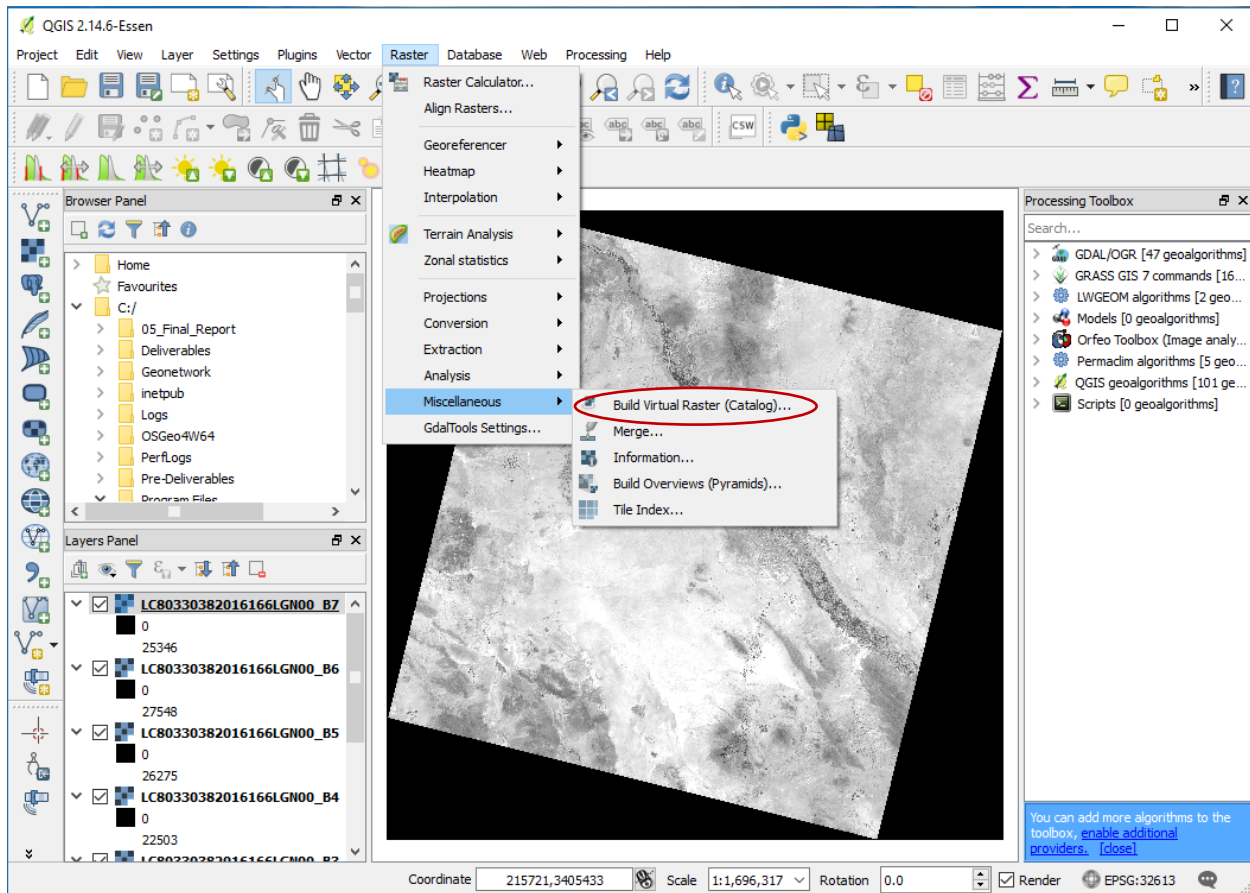


Once the Layers are added, double click any of them to access general information for each image. Note the spatial reference information: EPSG:32613, WGS 84 / UTM zone 13N).

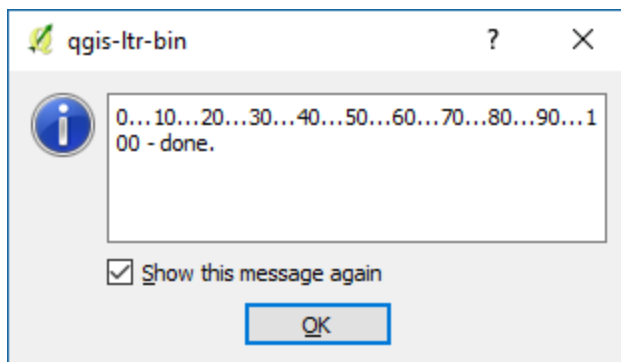
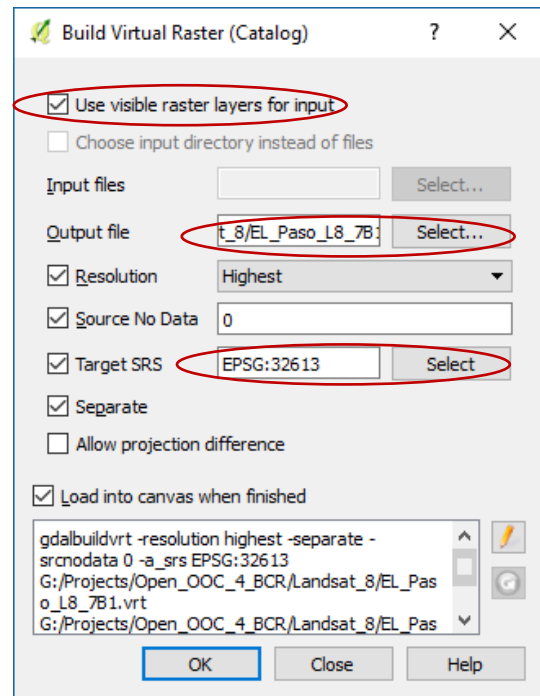


On the QGIS top menu, select **Raster > Miscellaneous > Build Virtual Raster (Catalog)**.

## Instructions for Open Source Object Oriented Classification



Under the **Build Virtual Raster (Catalog)** options, use the visible rasters for input, type a name for the output file (El\_Paso\_L8\_7B), check resolution to the highest, the source for no data as zero and to load into the canvas when finished by clicking the checkmarks. For the selection of the target Spatial Reference System (SRS), click the "Select" button, and under filter type the EPSG number (32613), then select: "EPSG:32613, WGS 84 / UTM zone 13N" as your SRS. Click **OK** for the SRS selection and Separate also, click **OK** to build your catalog. The process should take just a few seconds. Click OK.



Double click your layer stack (El\_Paso\_L8\_7B, new image with all 7 bands). You should be able to change the band rendering by right click and go to properties (to see the three bands that you can see at any given time). The “rendering type” should be Multiband color. For a natural color approximation choose red: band 4, green: band 3, blue: band 2. Under Contrast enhancement, you may want to select from the pull-down menu “Stretch to MinMax”. Click OK. A description of the 7 bands of your stack is below, all of which are acquired at 30 m ground resolution.

<b>Landsat 8 Band Name</b>	<b>Bandwidth (μm)</b>
Band 1 Coastal	0.43 – 0.45
Band 2 Blue	0.45 – 0.51
Band 3 Green	0.53 – 0.59
Band 4 Red	0.64 – 0.67
Band 5 NIR	0.85 – 0.88
Band 6 SWIR 1	1.57 – 1.65
Band 7 SWIR 2	2.11 – 2.29

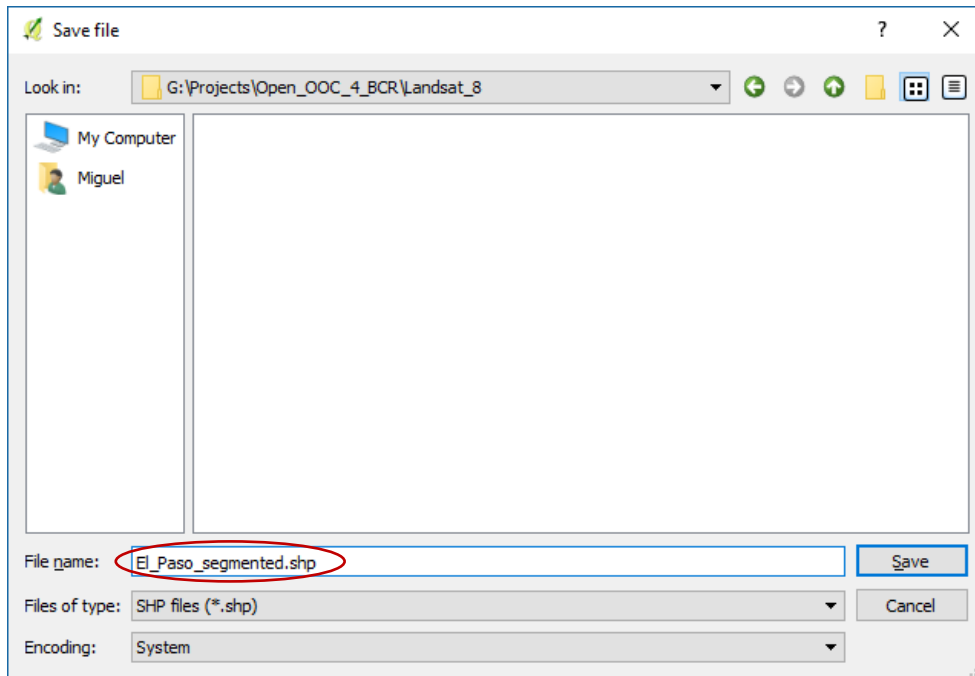
### 3.3 Vector Segmentation and Data Exploration

On the QGIS top menu, select **Processing > Toolbox**, to get access to the different modules on the Processing Toolbox, one of them should be **Orfeo Toolbox (Image analysis)**.

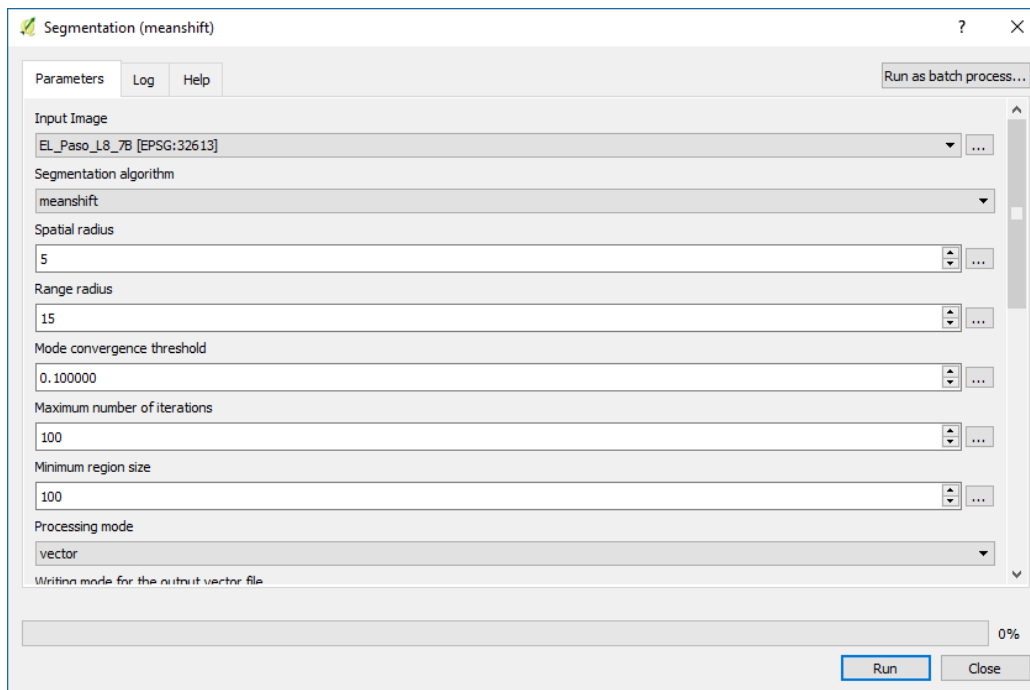
Expand the Orfeo Toolbox and then Segmentation toolset. There are several options for segmenting an image, the process of grouping contiguous pixels with similar color (spectral values) together in segments; for this exercise we will try the **Segmentation (meanshift)** option by double clicking on it.

The input image is “El\_Paso\_L8\_7B [EPSG:32613]”. We will keep the default values as they are. We will define a shapefile as the Output vector file, something meaningful like: “**El\_Paso\_segmented.shp**”. Click on the three periods to the right of the Output vector file field, then save to file, navigate to your directory, and type in “**El\_Paso\_segmented.shp**”.

## Instructions for Open Source Object Oriented Classification



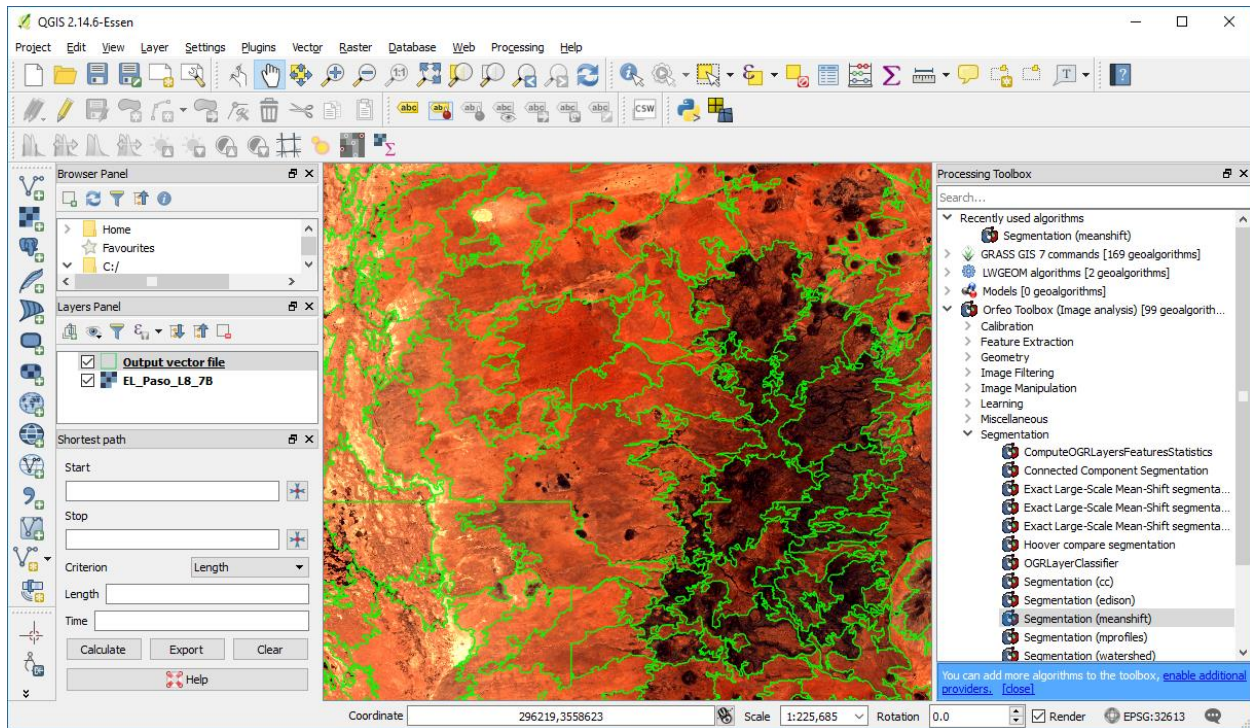
The segmentation tool should look something like this:



Click the **Run** button and let it process the image, it should take several minutes.

Depending on the level of resolution and the number of classes you may want to achieve, you may play with different combination on all the parameters on the meanshift segmentation until you get satisfactory results. Below is the result of a previous run with **Range radius** moved from 15 to 30 and **Minimum regions size** from 100 to 9000 after zooming in and double-clicking on the output layer, clicking on Style, clicking on Fill, and selecting "Transparent fill".

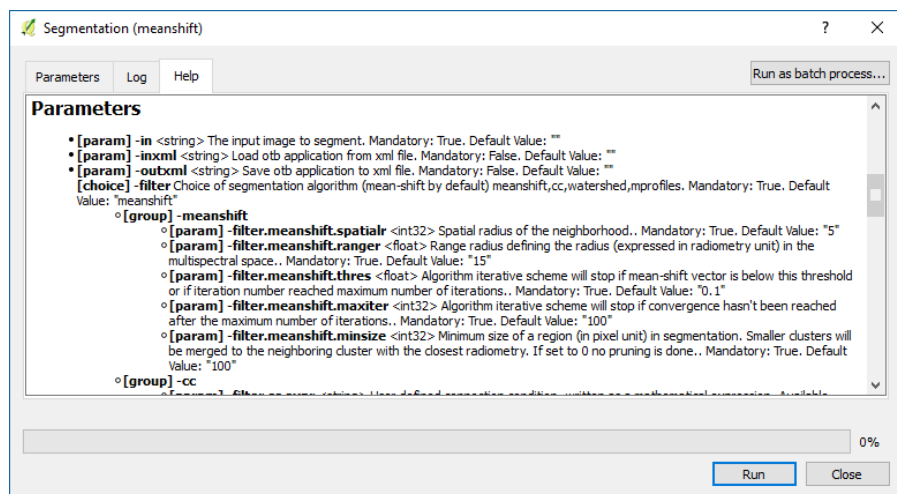
## Instructions for Open Source Object Oriented Classification



In the previous example you can see that there are still dark spots aggregated with the lighter colored background. This means the radius needs to be reduced and also the minimum regions size needs to be smaller, but now we have explored what happens if you provide big numbers for those parameters, you get segments that are too big and general.

Open the Segmentation (meanshift) tool again and move back the **Range radius** to 15 (default value) and region size to 100 (also default value). OTB looks for neighboring pixels whose spatial location (in number of pixels) falls within the spatial radius and whose Euclidean distance (expressed in radiometric units) between spectral values falls within the range radius.

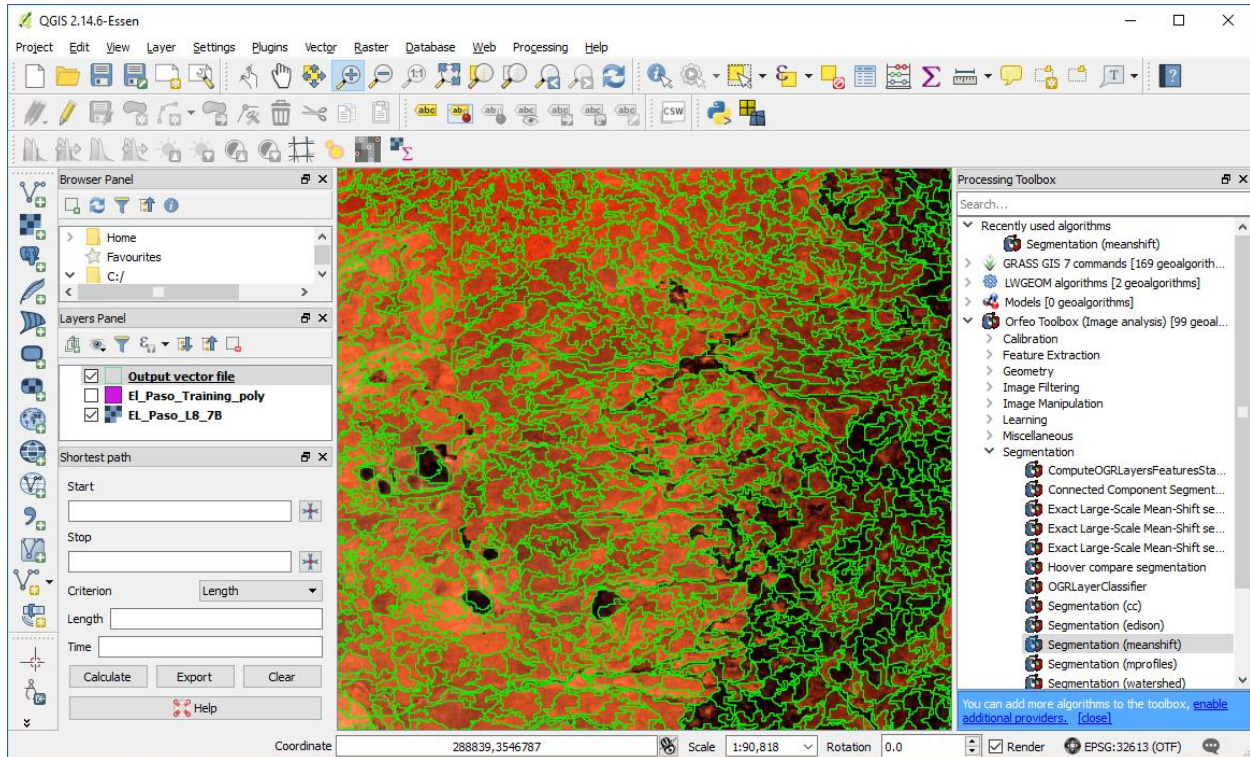
Explanation of what the tool does and what this parameters mean can be accessed by clicking on the "Help" tab provided with the tool. Zoom in near the center of the previous screen capture to see most of the darker spots now belong to just one segment as desired.






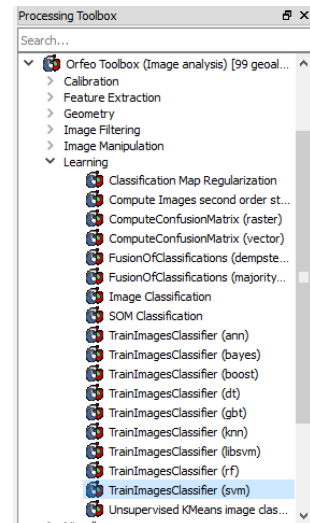
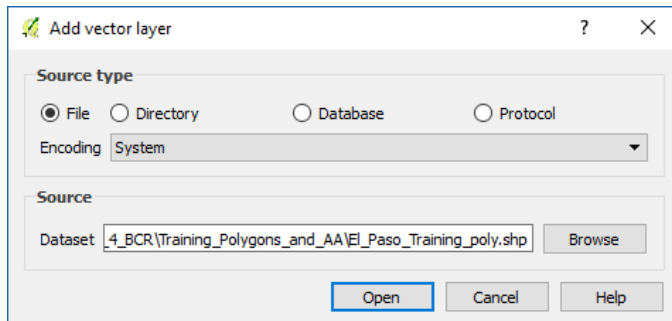
## Instructions for Open Source Object Oriented Classification

Below is the result of default values with Range radius back at 15 (default value) and region size at 100 (also default value).



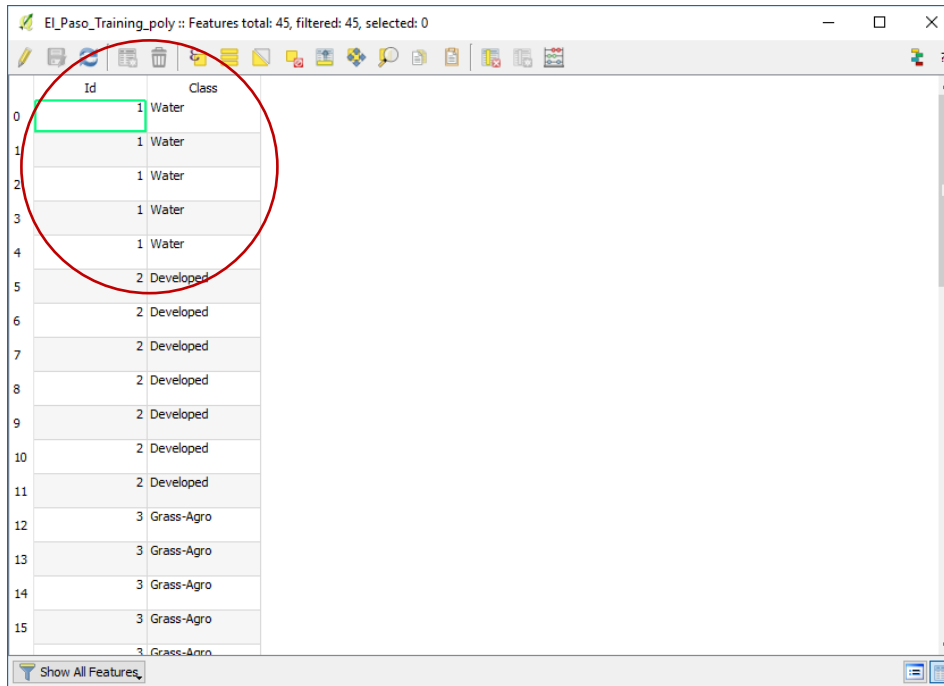
Once we have a segmentation that fits our needs, we need to train a classifier, and algorithm that does the classification work assigning segments to classes, it will use information from the Landsat image ("El\_Paso\_L8\_7B.vrt") and the training polygons vector data ("El\_Paso\_Training\_poly.shp").

On the left hand side of your QGIS window, you can see several icons with a plus sign. Those refer to datasets that you can add to your session. Hover over the first one  until you see a label called "Add Vector Layer", click on it. A dialog for adding a vector layer is opened, select the source type to be a "file", and under source, browse and navigate to where you have your training polygons, in this example we are using: "El\_Paso\_Training\_poly.shp". Click Open.



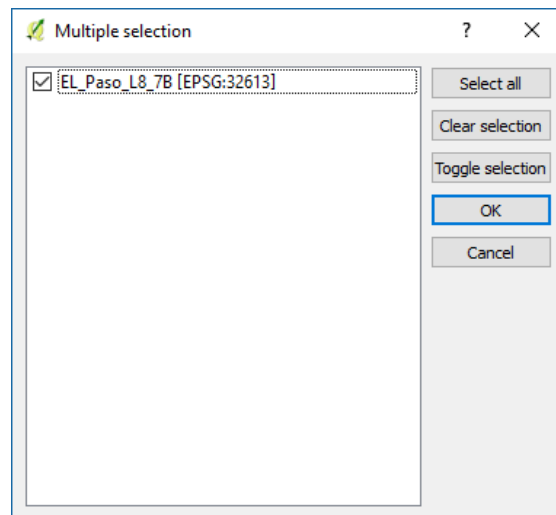
## Instructions for Open Source Object Oriented Classification

The training vector data must contain polygons with a positive integer field representing the class label. You can right-click under your Layers Panel: “El Paso Training poly” dataset and click “Open attribute table”. The table is opened. In this example we used 1 for water, 2 for developed, 3 for grass and agriculture, 4 for brushland, 5 for barren land and 6 for forest. The first value is under a field called “Id” and the second is under a field called “class”. Those can be seen in the table for “El\_Paso\_Training\_poly.shp” below.



	Id	Class
0	1	Water
1	1	Water
2	1	Water
3	1	Water
4	1	Water
5	2	Developed
6	2	Developed
7	2	Developed
8	2	Developed
9	2	Developed
10	2	Developed
11	2	Developed
12	3	Grass-Agro
13	3	Grass-Agro
14	3	Grass-Agro
15	3	Grass-Agro

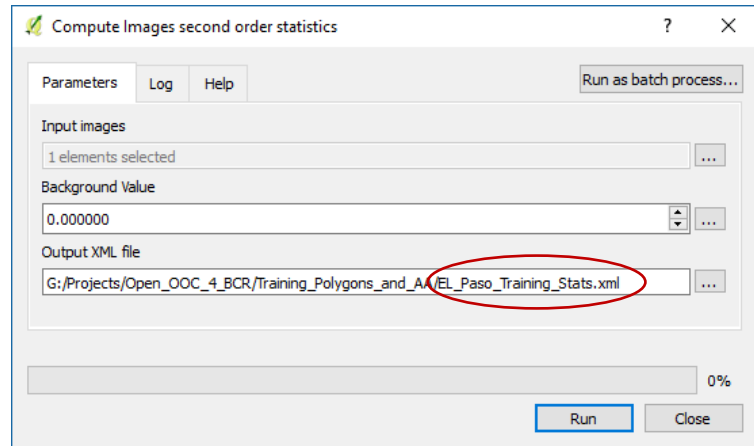
To get an idea of our image data contents, we need to calculate global mean and standard deviation for each band of the Landsat image (“El\_Paso\_L8\_7B.vrt”) and save the results in an XML file.





## Instructions for Open Source Object Oriented Classification

On the processing toolbox panel go to **Orfeo Toolbox > Learning > Compute Images second order statistics** and double click on it. As input we use “El\_Paso\_L8\_7B.vrt” (click on the three dots to select it). The background value is zero and we define an output xml file (click on the three dots and save to file): “El\_Paso\_Training\_Stats.xml”. Hit Run.

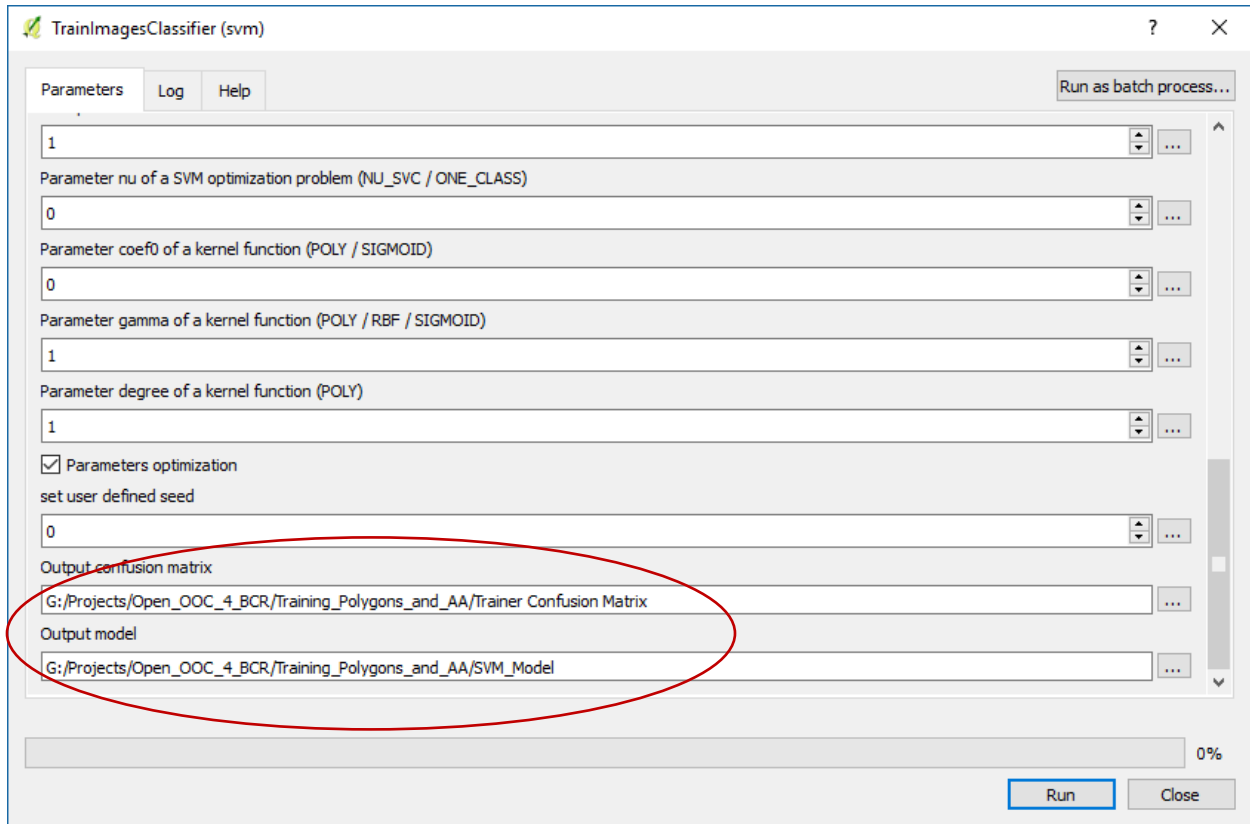


That calculates the global mean and standard deviation for our image and stores it in the “El\_Paso\_Training\_Stats.xml” file. This file can be used optionally by the TrainerImagesClassifier (svm) algorithm, or it can generate the statistics itself.

On the processing toolbox panel go to **Orfeo Toolbox > Learning > TrainImagesClassifier (svm)** and double click on it. This tool performs a classifier training from an image(s) and training vector data. The training vector data must contain polygons with a positive integer field representing the class label (which is “Id” field in our case).

Check the image and the training polygons files that will be used, we are using the Landsat image (“El\_Paso\_L8\_7B.vrt”) as the **Input Image** and training polygons vector data (“El\_Paso\_Training\_poly.shp”) as the **Input Vector Data**. We will use most of the default values in the **TrainImagesClassifier (svm)** tool, except for the **Name of the Discriminating field** (an integer field) we will type “Id”. Scrolling to near the end of the dialog, you will find a place to specify the output confusion matrix and the output model. The first one will give us an idea of how similar and discernible are the classes among themselves. The more values that land on the diagonal, the least confusion (call the output file something meaningful like “Trainer Confusion Matrix”). The second will give us the model needed to perform the classification (call it “SVM\_Model”) in the following step. A support vector machine (SVM) Model is a supervised machine-learning model with associated learning algorithms in QGIS that analyze the data used for the object-oriented classification in this case.

## Instructions for Open Source Object Oriented Classification



Click **run** to start the TrainImageClassifier tool.

The Training Image classifier produces a confusion matrix in text format, an extension ".txt" can be added to it to make it recognizable by text editors such as notepad. Once it can be open, it will give us an idea of how classes are potentially classified with the training polygons used. It can also be imported into a spreadsheet and reformatted to include the class names. The confidence levels of all the class labels are estimated from a comparison of the classification maps to ground truthed classes, which results in a confusion matrix. The reference labels for each class are rows, one per class defined in the training polygons. The Produced labels are columns, one per class in the resulting classification. An acceptable confusion matrix will have most of the highest values in the diagonal, meaning that the classification resembles what has been observed on the field, for example, 128 pixels referenced as Water during ground truthing were correctly classified as Water in the matrix below.

Values that are not in the diagonal are confused. For example there are 5 pixels referenced as Developed in the matrix below that were misclassified as Water. These can be asphalt or other land cover that the classifier is confusing with Water, therefore the name, confusion matrix. For each classification map, these confusion matrices are then used to estimate the "mass of belief" of each class label. The more values that land in the diagonal, the more belief in the classifier doing a good job for a given class. If the discernibility is poor, that shows as many pixels not in the diagonal; more ground truthing or re-selection of the training polygons used may be needed. This is an example of what can be expected as the Confusion matrix, which gives an idea of the discernibility of the classes:

## Instructions for Open Source Object Oriented Classification

#Produced labels (columns)	1	2	3	4	5	6
#Reference labels (rows)						
1 Water	128	3	4	5	0	2
2 Developed	5	128	0	5	0	0
3 Grass-Agro	6	0	144	2	0	1
4 Brushland	1	2	0	151	0	0
5 Barren land	0	0	0	6	159	0
6 Forest	0	1	1	0	0	138

Later in the process a confusion matrix will be used as the accuracy assessment is performed by comparing the map created by the Object-Oriented Classification against the Ground-Truth resources available.

### 3.4 Raster Segmentation and Object-Oriented Classification

Raster segmentation is the process of partitioning a digital image into multiple segments, sets of pixels or objects. The goal of segmentation is to pre-process or simplify the representation of an image into objects that are more meaningful and easier to analyze. Pixels that are spatially close and are spectrally similar are grouped into objects. With the OTB segmentation and classification tools, users can prepare segmented rasters as a first step in creating object-oriented classified raster datasets.

Now that we have a better understanding of the Landsat 8 image, training polygons, and their classes, a raster segmentation is needed for the object-oriented classification. Each processing step may take several minutes depending on the amount of data to process and the computer speed. The workflow in QGIS comes in three main procedures:

- Mean-shift raster segmentation of large images (4 steps).
- Preparation of reference data (training polygons).
- Object Oriented Classification using the SVM algorithm.

#### 3.4.1 Mean-shift segmentation workflow

- a. Perform a mean-shift filtering of "El\_Paso\_L8\_7B.vrt". This will help extract homogeneous objects from a filtered image. On the processing toolbox panel go to **Orfeo Toolbox > Image Filtering > Exact Large-Scale Mean-Shift segmentation, step 1 (Smoothing)**.

## Instructions for Open Source Object Oriented Classification

Exact Large-Scale Mean-Shift segmentation, step 1 (smoothing)

Parameters Log Help Run as batch process...

Input Image  
EL\_Paso\_L8\_7B [EPSG:32613]

Available RAM (Mb)  
10000

Spatial radius  
5

Range radius  
30

Mode convergence threshold  
0.100000

Maximum number of iterations  
100

Range radius coefficient  
0

☒ Mode search.

Filtered output  
El Paso L8 Filtered

☒ Open output file after running algorithm

Spatial image  
El Paso L8 Spatial

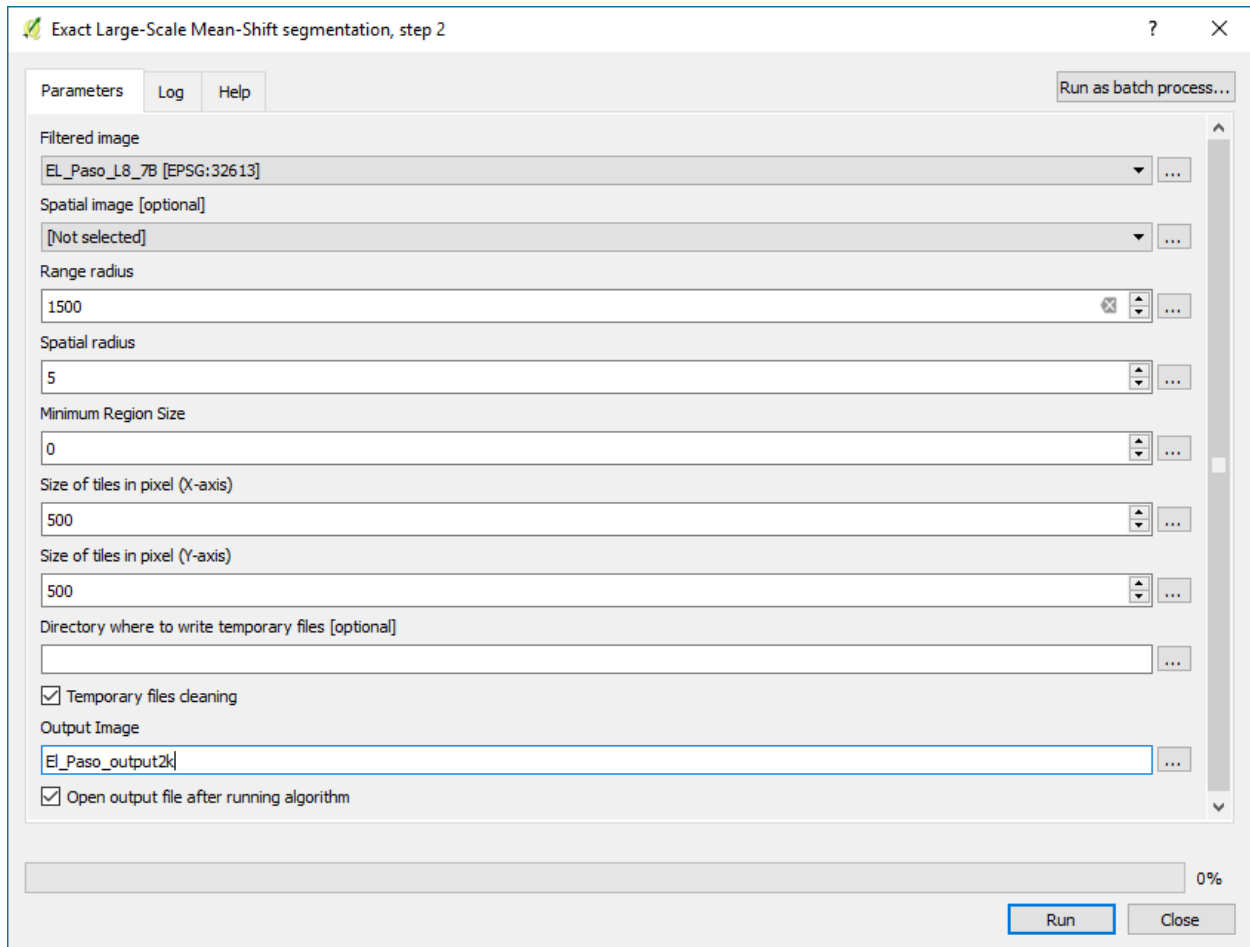
☒ Open output file after running algorithm

0%

Run Close

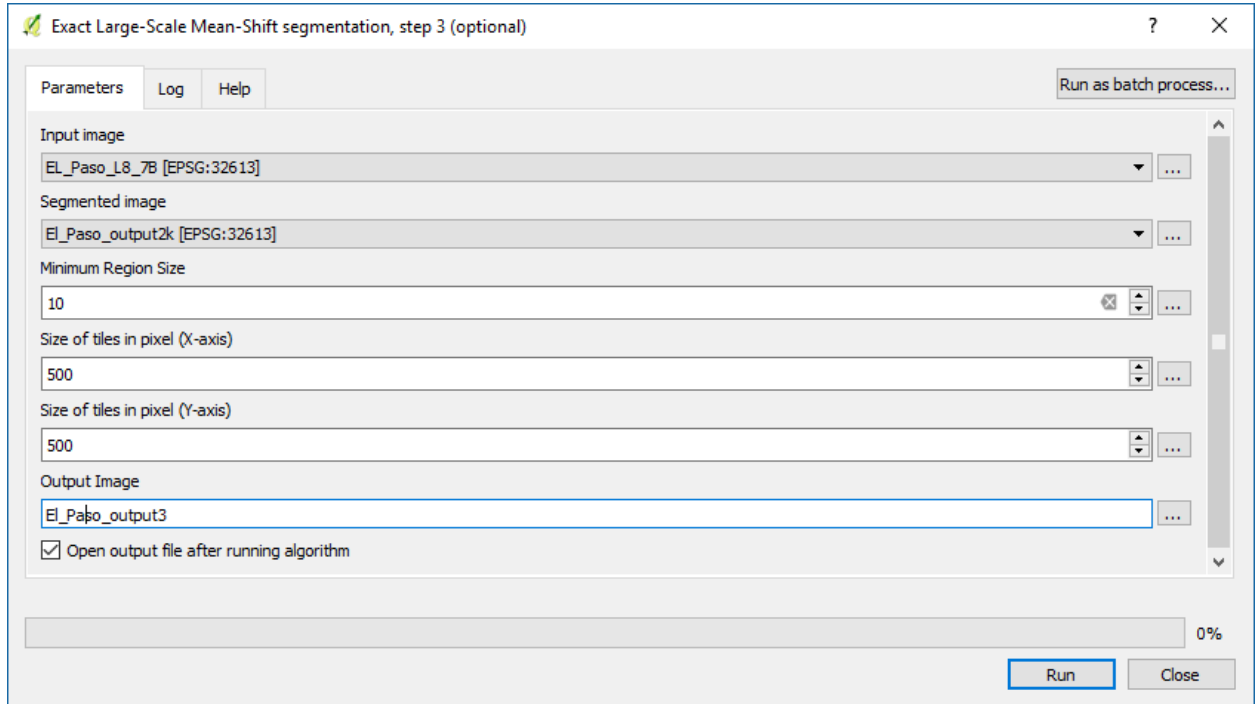
- Use "El\_Paso\_L8\_7B.vrt" as the **input image**.
- Set the **Available RAM (Mb)** to what you want to make available to GIS to process. The more RAM your computer has available, the faster the process runs.
- Set the **Range Radius** to 30.
- Click on the three dots and select "save to file" to define a path for "**filtered output**" (name the file "El\_Paso\_L8\_filtered.tif") and **spatial image** (name it "El\_Paso\_L8\_spatial.tif").
- Make **Mode Search** parameter disabled.
- Leave all the other parameters with the default values and click **run**.
- Notice that "El\_Paso\_L8\_filtered.tif" has less variance than the original image, it looks smoother. Some of the heads and tails on the original have been filtered resulting on a smoother image. This can be verified by right-clicking on the image properties, click on Histogram, and calculate the histogram for each layer. You can capture the screen of both histograms as a reference, the filtered image has more pixels near the center of the distributions and less towards the extremes (heads and tails of the distribution curve).

- b. Convert the filtered image to a one-band object. This step will produce a labeled image where neighbor pixels whose range distance is below range radius (and optionally spatial distance below spatial radius) will be grouped together into the same cluster. Go to **Processing-Toolbox --> Orfeo Toolbox (image analysis) --> Segmentation --> Exact Large-Scale Mean-Shift segmentation, step 2.**



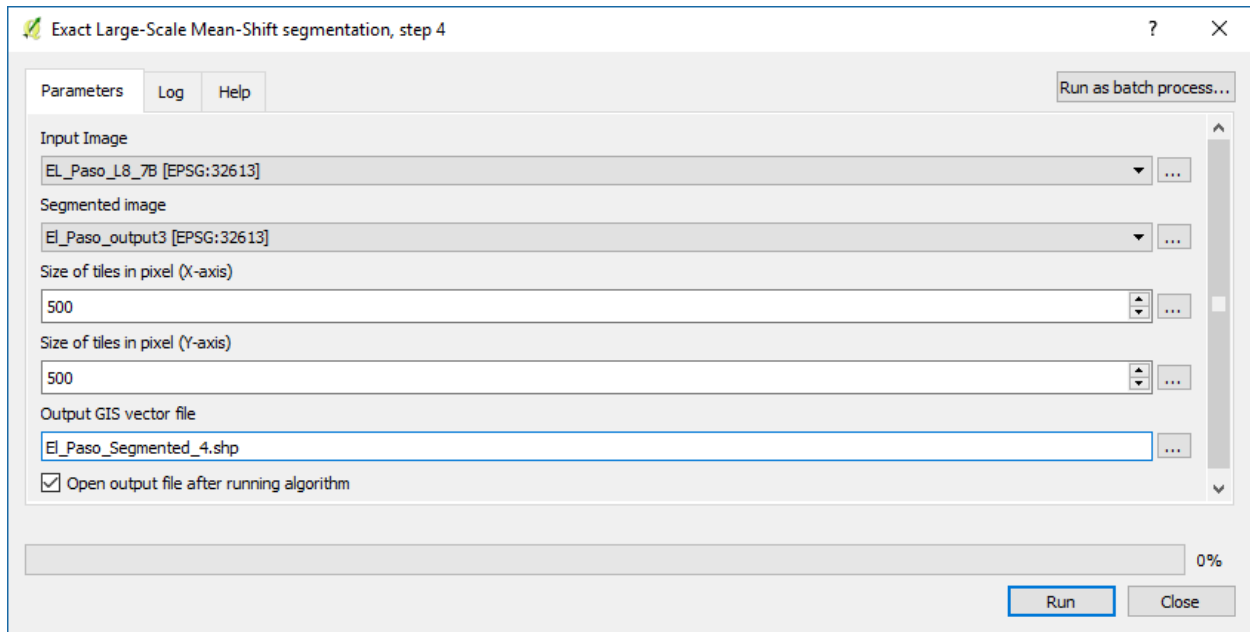
- i. Use “EL\_Paso\_L8\_filtered.tif” as **Filtered Image** (output from previous step).
- ii. Leave the **Spatial image (optional)** empty (not selected).
- iii. Increase the **Range Radius** to “1500”
- iv. Click on the three dots and select “save to file” to save the **Output Image** as a file name meaningful to you; in this example we use “El\_Paso\_L8\_output2k.tif”.
- v. Leave all the other parameters with the default values and click **run**.
- vi. Notice the groupings of pixels into objects in the output image (“El\_Paso\_L8\_output2k”). You can change the **Range Radius** to other values. When distinct features on the ground (i.e. the dark spots mentioned earlier), are in segments by themselves, you know that you have good parameters, continuous features on the ground broken excessively into small segments, or big segments covering several features or classes on the ground need to be refined accordingly. Repeat as necessary to satisfy your classification needs if the initial results are not satisfactory.

- c. Merge small objects whose size in pixels is lower than **minsize** parameter with the adjacent region that has close Euclidian radiometric distance and acceptable size. Go to **Processing-Orfeo Toolbox (image analysis) --> Segmentation --> Exact Large-Scale Mean-Shift segmentation, step 3 (optional)**.

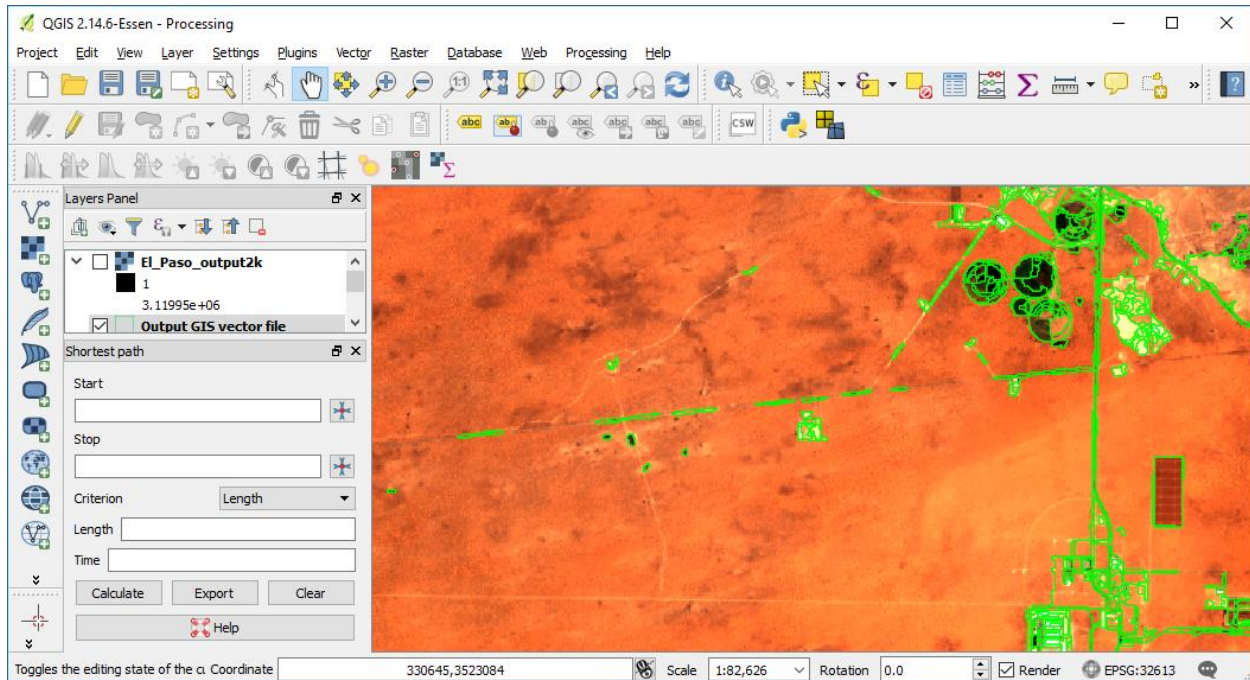


- i. The **Input image** is the original image stack ("EL\_Paso\_L8\_7B").
  - ii. The **Segmented image** ("El\_Paso\_output2k.tif") is the result of the raster segmentation above (step 1b).
  - iii. The **Minimum Region size** is defined as 10.
  - iv. Click on the three dots and select "save to file" to save the **Output Image** as a file name meaningful to you. In this example we use "El\_Paso\_output3.tif".
  - v. Leave all the other parameters with the default values and click **run**.
  - vi. You may notice that small objects have been absorbed by bigger, contiguous objects that are close radiometrically (an Euclidian distance in digital values that is within a pre-defined tolerance).
- d. Transform the raster image objects to polygon (vector objects). Each polygon contains additional fields: mean and variance of each channel from input image (in parameter), segmentation image label, and number of pixels in the polygon. Go to **Processing-Orfeo Toolbox (image analysis) --> Segmentation-Exact Large-Scale Mean-Shift segmentation, step 4**.

## Instructions for Open Source Object Oriented Classification



- Use “El\_Paso\_L8\_7B.vrt”. as the **Input image**.
- Use “El\_Paso\_output3.tif” as the **Segmented Image**.
- Click on the three dots, select “save to file” and name the **Output GIS vector file**. In this example “El\_Paso\_Segmented\_4.shp” is used.
- Leave all the other parameters with the default values and click **run**.



- Notice that now we have vector objects that have mean and variance statistics for each of the 7 bands in the input images (0-6). Right click on the **Output GIS vector file** in your QGIS session (“El\_Paso\_Segmented\_4.shp” in

## Instructions for Open Source Object Oriented Classification

reality) and click on **Open Attribute Table**. Verify that the values for label, nbPixels, means, and variances have been populated for every band (column).

Output GIS vector file :: Features total: 192678, filtered: 192678, selected: 0

	label	nbPixels	meanB0	meanB1	meanB2	meanB3	meanB4	meanB5
0	1	19718168	0.094608888030...	0.275439739227...	0.330502420663...	0.238587483763...	0.551012992858...	1.149394750
1	4	4103039	12602.39160156...	12572.40234375...	13067.26269531...	15758.64257812...	19901.46679687...	22304.12695
2	93	24	13466.79199218...	13455.62500000...	13941.29199218...	16065.54199218...	19210.54101562...	23482.75000
3	85	20	13397.75000000...	13365.75000000...	13734.15039062...	15657.54980468...	18769.05078125...	23712.80078
4	103	15	13501.79980468...	13512.86621093...	14064.73339843...	16265.46679687...	19482.06640625...	23667.19921
5	83	52	13616.05761718...	13660.50000000...	14247.50000000...	16521.90429687...	19873.71093750...	24201.30859
6	90	36	14015.52734375...	14145.27734375...	14941.55566406...	17488.97265625...	21064.86132812...	25007.13867
7	331	14	14348.28613281...	14562.85742187...	15530.64257812...	18408.28515625...	22183.57226562...	24282.57226
8	495	18	13097.22265625...	13005.77734375...	13243.72265625...	15007.83300781...	18390.38867187...	22302.33398
9	732	25	13546.75976562...	13596.59960937...	14216.04003906...	16417.11914062...	20588.88085937...	23189.35937
10	852	12	12210.08300781...	11949.83300781...	12195.58300781...	13315.33300781...	20328.50000000...	19670.58398
11	1129	13	12309.46191406...	11958.46191406...	11835.61523437...	12605.92285156...	17257.46093750...	17260.46093
12	1260	12	13577.58300781...	13529.16699218...	13865.66699218...	15815.83300781...	19057.58398437...	20733.41601
13	1365	17	14553.76464843...	14677.52929687...	15295.35253906...	17505.64648437...	19836.64648437...	20095.47070
14	1261	193	13952.73046875...	14034.22265625...	14732.34179687...	17060.87500000...	20674.81250000...	22896.20703
15	1400	31	14671.54882812...	14775.25781250...	15427.48339843...	17806.19335937...	21565.54882812...	21854.87109

Show All Features

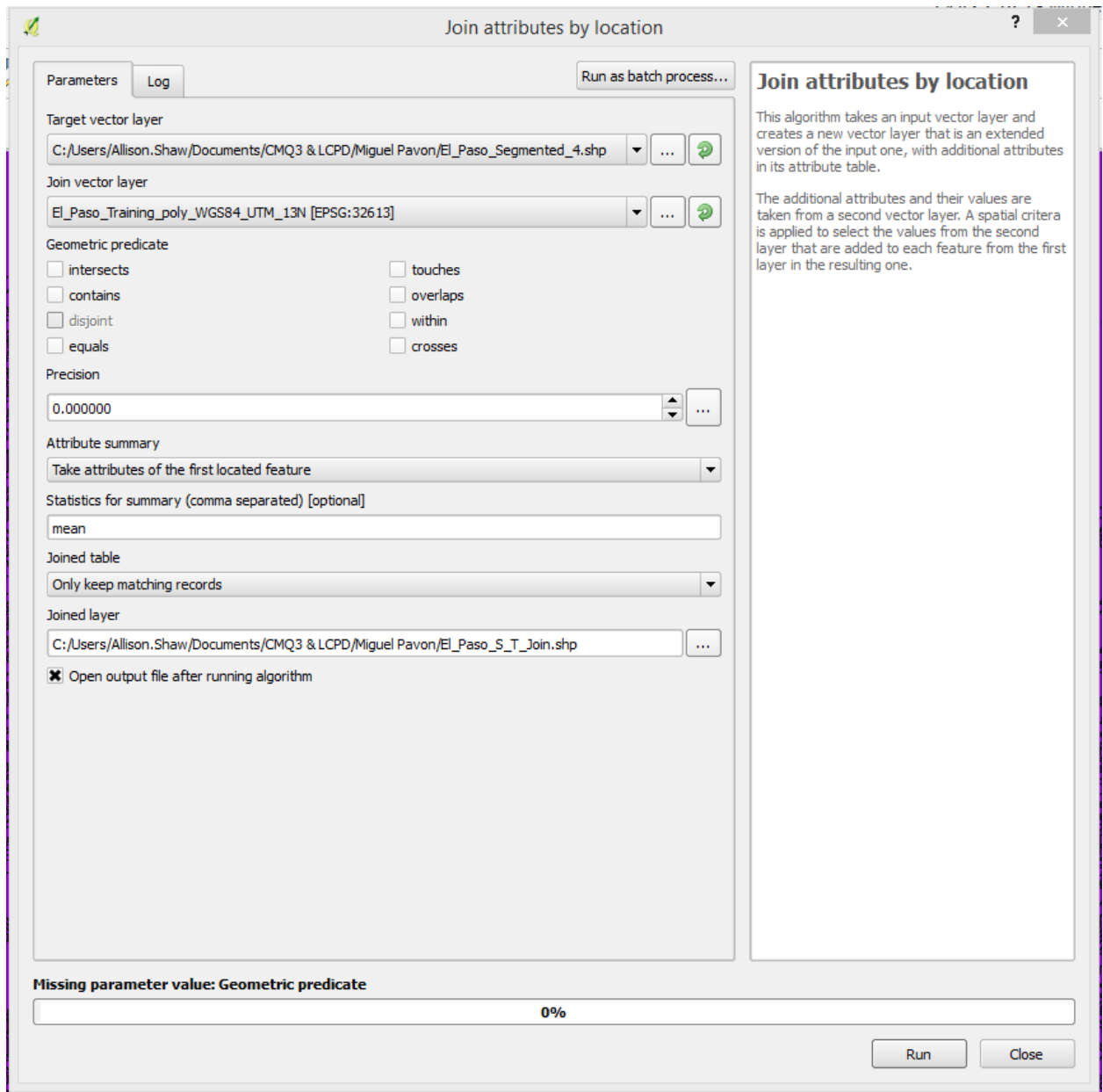
### 3.4.2 Preparation of reference data

For this step we will use training polygons vector data ("El\_Paso\_Training\_poly.shp") created as sample polygons and used before during data exploration.

- Join the training polygons class attribute with the image segments. Add vector dataset "El\_Paso\_Training\_poly.shp" to your QGIS Session (**Add Vector Layer** Icon in the left bar).
- Go to the main menu in QGIS (top bar) **Vector --> Data Management Tools --> Join Attributes by Location**.
  - The **Target vector file** is the Output of step 1d ("El\_Paso\_Segmented\_4.shp").
  - Set the **Join vector layer** as "El\_Paso\_Training\_poly.shp".
  - Click on the three dots for **Joined Layer**, select "Save to File" and save as "El\_Paso\_S\_T\_Join.shp".
  - Leave the rest of the parameters with the default values and click **Run**.



## Instructions for Open Source Object Oriented Classification



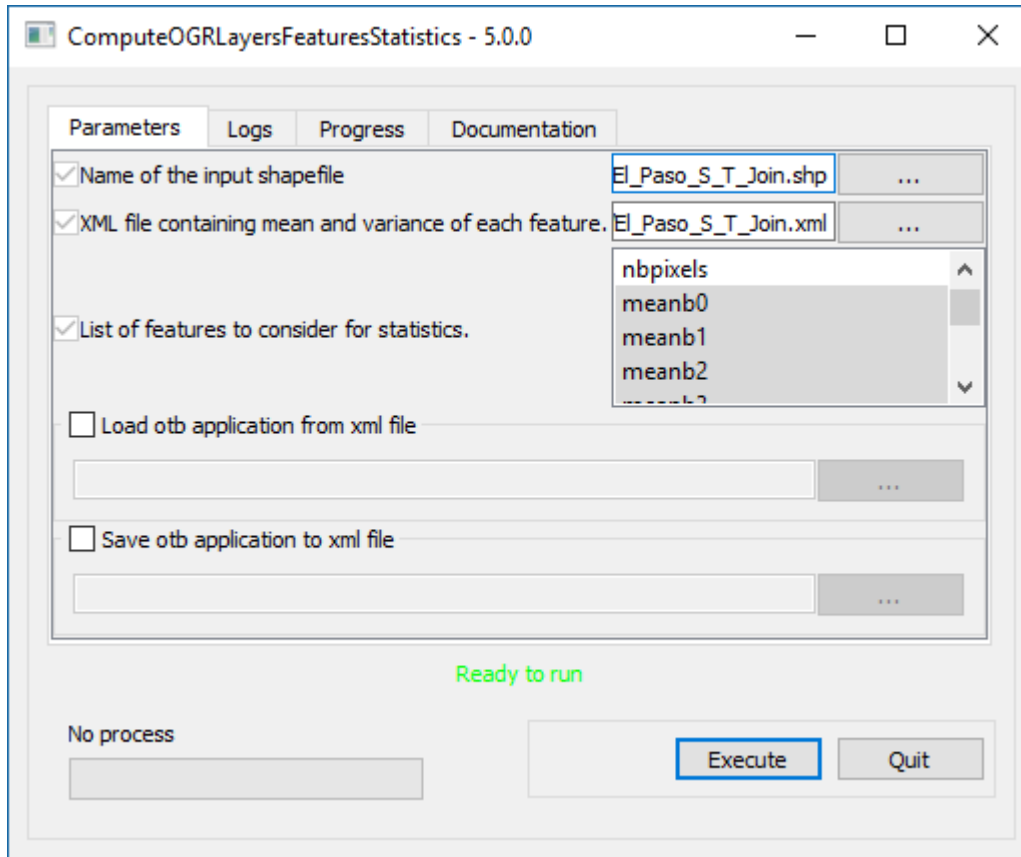
### 3.4.3 Object-based classification using SVM algorithm

- Open the Orfeo Tool Box graphical Interface by searching for the OSGeo4W Shell. In Windows you can find it here: **Start --> All Programs --> OSGeo4W --> OSGeo4W Shell.**

```
OSGeo4W Shell - otbgui_ComputeOGRLayersFeaturesStatistics
07/30/2016 07:52 AM 459,511 Pavon_signed.pdf
11/14/2016 11:30 PM 25,661 Processing.qgs
11/13/2016 06:13 PM <DIR> Training_Polygons_and_AA
5 File(s) 675,413 bytes
5 Dir(s) 3,249,401,733,120 bytes free

G:\Projects\Open_OOC_4_BCR>cd Landsat*
G:\Projects\Open_OOC_4_BCR\Landsat_8>cd prueba
G:\Projects\Open_OOC_4_BCR\Landsat_8\prueba>otbgui_ComputeOGRLayersFeaturesStatistics
```

- b. On the command shell, type: **otbgui\_ComputeOGRLayersFeaturesStatistics**, to open a new Graphic User Interface (GUI).



- i. Click on the three dots by **Name of the input file** and navigate to "El\_Paso\_S\_T\_Join.shp".
- ii. Click on the three dots by **XML file containing mean and variance of each feature** and type in "El\_Paso\_S\_T\_Join.xml".
- iii. Select from the **List of features to consider for statistics**: meanb0, meanb1, meanb2, meanb3, meanb4 meanb5 and meanb6.
- iv. Click: **Execute**.
- v. Notice a new xml file ("El\_Paso\_S\_T\_Join.xml") was created with contents defining the mean and standard deviation of each column considered in step 3.b.iii:

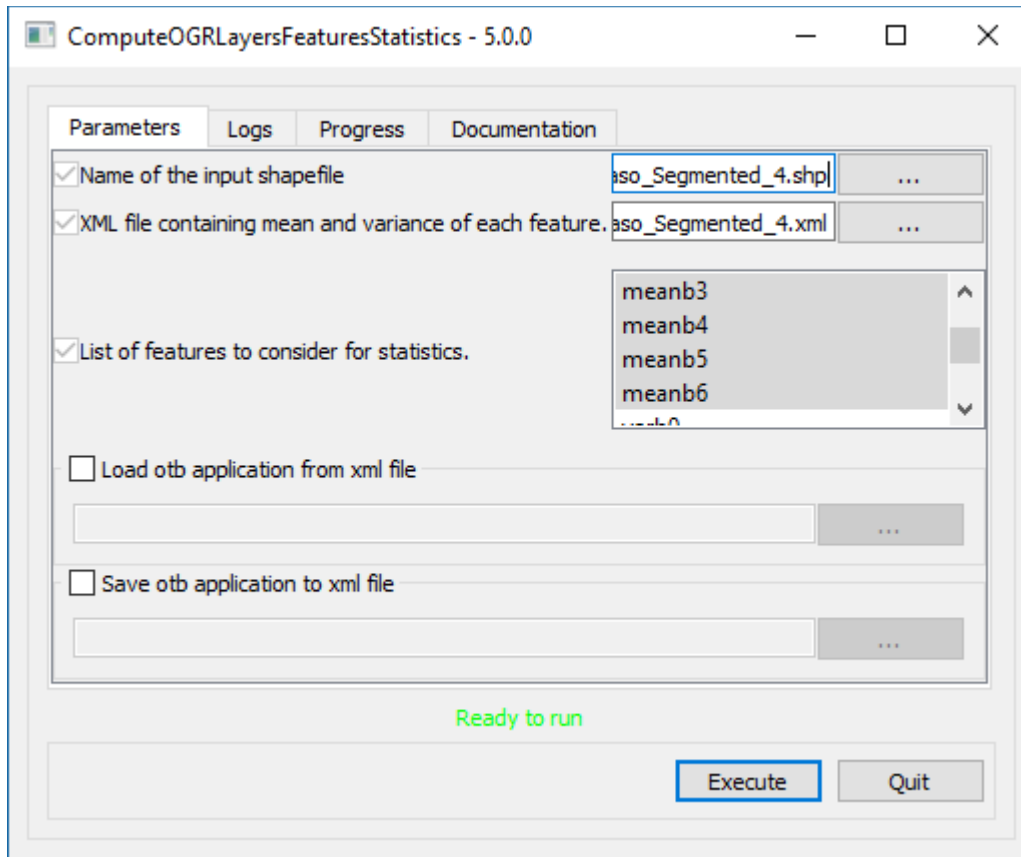
```
<?xml version="1.0"?>
```

```
<FeatureStatistics><Statistic name="mean"><StatisticVector
value="12233.5"/><StatisticVector value="11915.7"/><StatisticVector
value="12262"/><StatisticVector value="13163.8"/><StatisticVector
value="20137.1"/><StatisticVector value="18113"/><StatisticVector
value="15200.3"/></Statistic><Statistic name="stddev"><StatisticVector
value="1676.7"/><StatisticVector value="1970.47"/><StatisticVector
value="2305.63"/><StatisticVector value="3251.66"/><StatisticVector
```

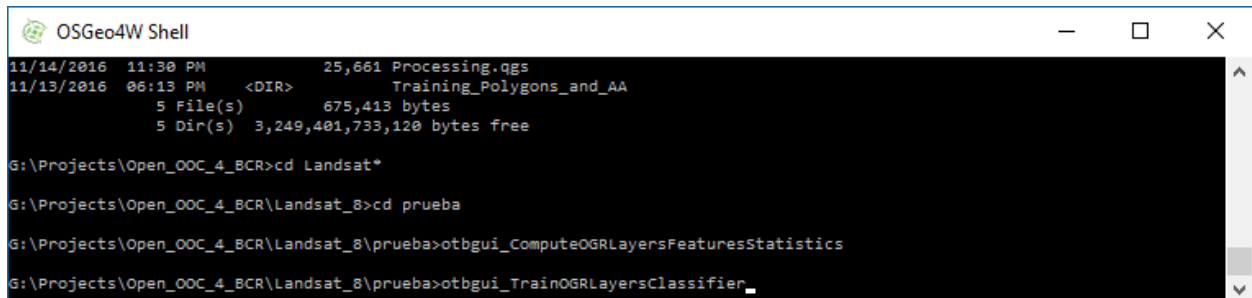
## Instructions for Open Source Object Oriented Classification

```
value="4175.49"/><StatisticVector value="3956.02"/><StatisticVector  
value="4072.67"/></Statistic></FeatureStatistics>
```

- vi. Repeat steps i.-iv. but changing “Name of the input shapefile” to “El\_Paso\_Segmented\_4.shp” and “XML file containing mean and variance of each feature” to “El\_Paso\_Segmented\_4.xml” to calculate xml statistics using the same parameters as before. When the green text “Ready to run” reappears, click “**Execute**”, when done, click **Quit**.



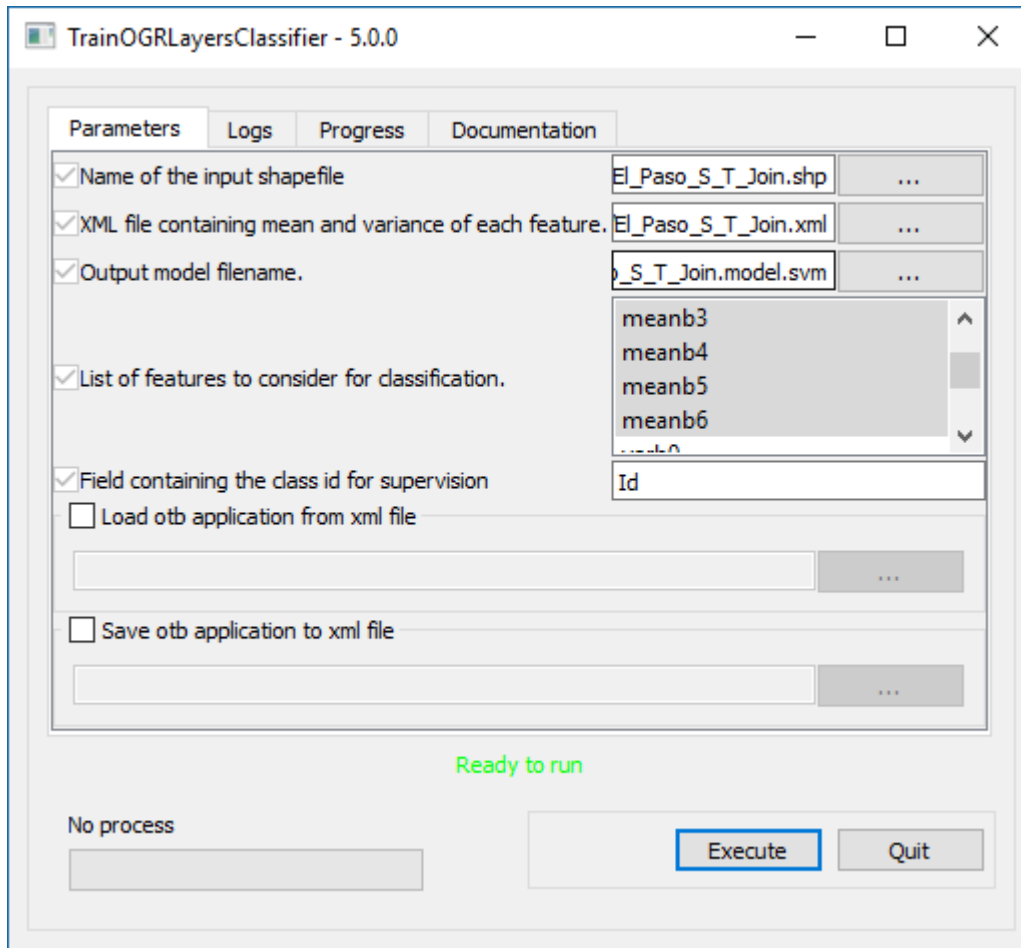
- c. On the command shell, type: **otbgui\_TrainOGRLayersClassifier**. A new GUI opens.



- i. Click on the three dots by **Name of input shapefile** and select “El\_Paso\_S\_T\_Join.shp”.

## Instructions for Open Source Object Oriented Classification

- ii. Click on the three dots by **XML file containing mean and variance of each feature** and select “El\_Paso\_S\_T\_Join.xml”.
- iii. Click on the three dots by **Output model name** and type: “El\_Paso\_S\_T\_Join.model.svm”.
- iv. Select from the **List of features to consider for classification**: meanb0, meanb1, meanb2, meanb3, meanb4, meanb5 and meanb6.
- v. On the **Field containing the class id for supervision**, type the field with the numeric value: “Id”.



- vi. Click: **Execute**. After finished, click **Quit**.
- vii. Notice there are notifications on the command shell and there is also a new model in your directory called “El\_Paso\_S\_T\_Join.model.svm” which can be opened with Notepad and should look something similar to:

## Instructions for Open Source Object Oriented Classification

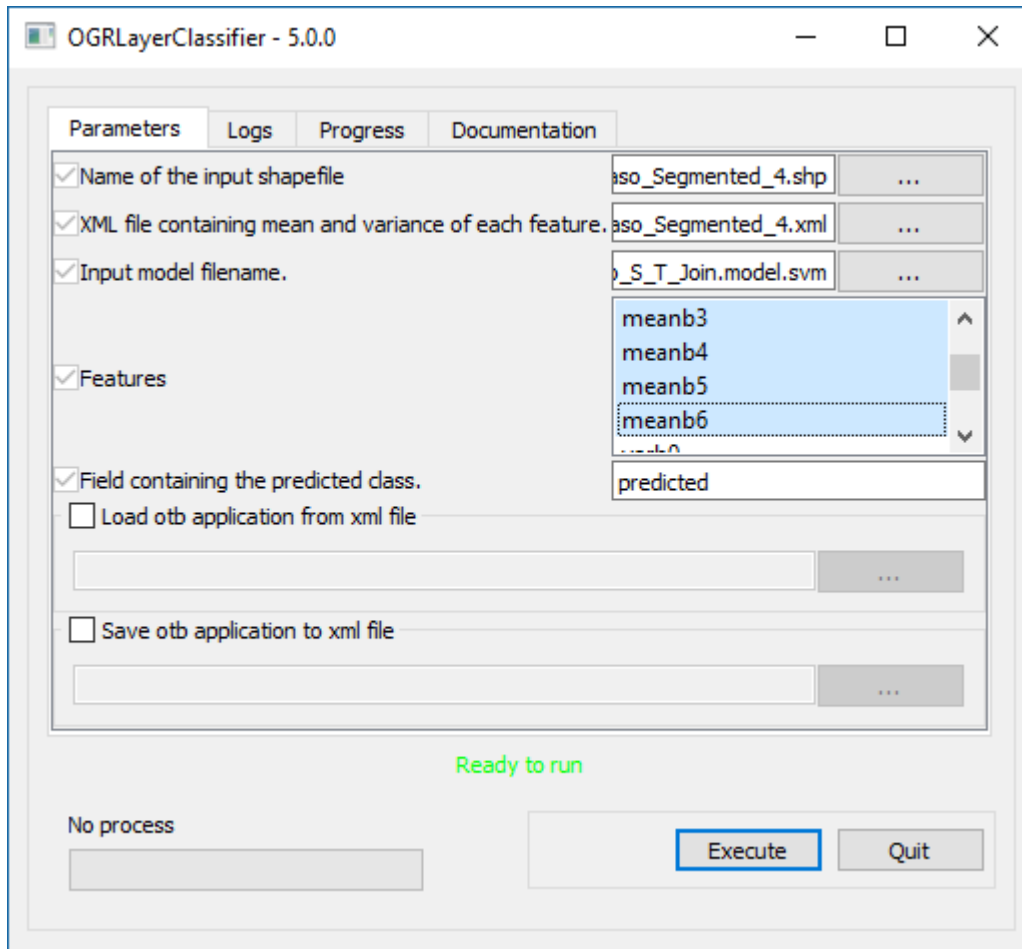
```
El_Paso_S_T_Join.model.svm - Notepad
File Edit Format View Help
svm_type c_svc
kernel_type linear
nr_class 6
total_sv 118
rho -1.91128 -3.09791 -2.9023 -1.0501 -0.574891 -0.44408 -3.56953 0.631438 -0.154755 -
1.06383 1.19069 0.513531 2.82539 0.496643 0.436799
label 3 4 6 5 1 2
nr_sv 30 27 10 6 16 29
SV
0.8705505632961241 0 0 0.8705505632961241 0.8705505632961241 1:-0.26961966 2:-0.25678449
3:-0.25279747 4:-0.067684239 5:-1.1318672 6:-0.74745039 7:-0.44360516
0 0.2881804030260093 0 0.8705505632961241 0.8705505632961241 1:-0.88775571 2:-0.9465078
3:-1.1070905 4:-0.91377522 5:-1.8883441 6:-2.0532819 7:-1.7007676
0.8705505632961241 0 0.8705505632961241 0.5155751139340982 0.8705505632961241 1:1.1308742
2:1.1746296 3:1.2501019 4:1.2692312 5:0.63154068 6:1.2275882 7:1.1586257
0 0 0 0.07549942109361806 1:-0.6438434 2:-0.41068277 3:-0.16256826 4:0.0057890503
5:0.9906125 6:-0.78761268 7:-1.0938821
0.8705505632961241 0 0.4872095764884074 0 0 1:0.38519266 2:0.76212147 3:1.2737028
4:1.7371412 5:1.2517907 6:2.271461 7:1.4416767
0.8705505632961241 0 0 0 1:0.076779001 2:0.16535613 3:0.23558713 4:0.35698317 5:-
```



- d. On the command shell, type: **otbgui\_OGRLayerClassifier**. A new GUI opens.

```
OSGeo4W Shell
optimization finished, #iter = 16
nu = 0.030075
obj = -2.401287, rho = 0.513531
nSV = 5, nBSV = 3
*
optimization finished, #iter = 7
nu = 0.120910
obj = -1.078415, rho = 2.825394
nSV = 4, nBSV = 1
*,*
optimization finished, #iter = 296
nu = 0.048472
obj = -3.925506, rho = 0.496643
nSV = 9, nBSV = 4
*,*
optimization finished, #iter = 224
nu = 0.176109
obj = -19.569696, rho = 0.436799
nSV = 27, nBSV = 22
Total nSV = 118
G:\Projects\Open_OOC_4_BCR\Landsat_8\prueba>otbgui_OGRLayerClassifier
```

This application will apply a trained machine learning model on the selected features to get a classification of each polygon contained in a layer. The list of features must match the list used for training. The predicted land cover class is written for each geometry into a user-defined attribute field which we have named "Predicted".

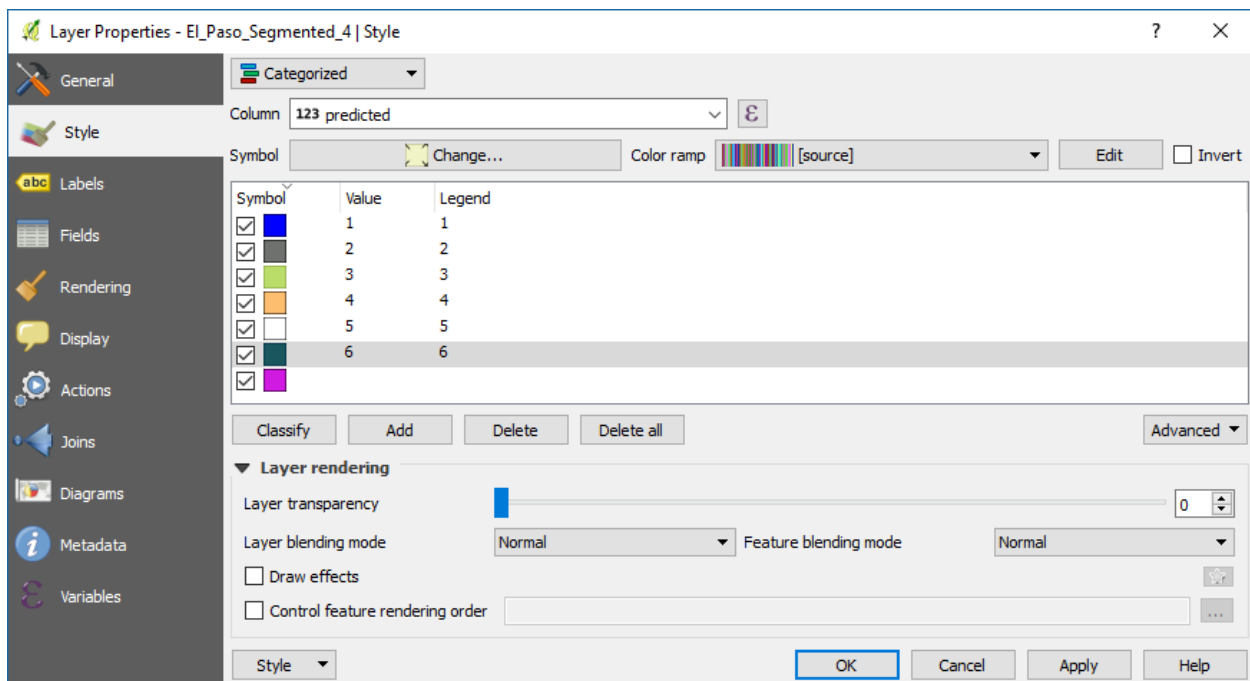




- i. Click on the three dots for **Name of input shapefile** and use the output of the segmentation: "El\_Paso\_Segmented\_4.shp".
  - ii. Click on the three dots by **XML file containing mean and variance of each feature** and select "El\_Paso\_Segmented\_4.xml".
  - iii. Click on the three dots by **Input model name** and use "El\_Paso\_S\_T\_Join.model.svm".
  - iv. Select from the list of **Features**: meanb0, meanb1, meanb2, meanb3, meanb4, meanb5 and meanb6.
  - v. Use "predicted" for **Field containing the predicted class**
  - vi. Click: **Execute**. It should take a couple of minutes, after finished, click **Quit**.
- e. Now you can open the resulting vector file in QGIS
- i. Open a new QGIS session.
  - ii. Add raster dataset "EL\_Paso\_L8\_7B" to your QGIS Session (**Add Raster Layer** Icon  in the left bar).
  - iii. Add vector dataset "El\_Paso\_Segmented\_4.shp" to your QGIS Session (**Add Vector Layer** Icon  in the left bar).

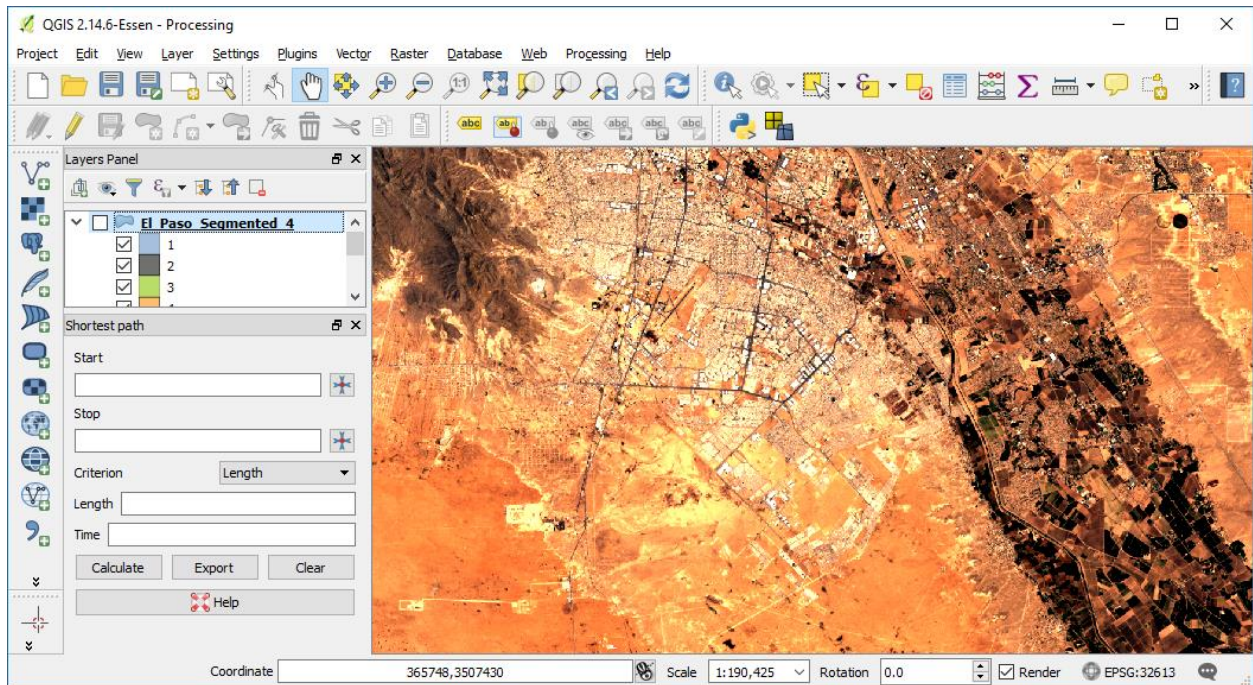
## Instructions for Open Source Object Oriented Classification

- iv. Right click “El\_Paso\_Segmented\_4” and select **Properties**. Click on the Style menu in the top left. Near the top, change from “Single symbol” to “Categorized.” For Column, scroll down to select “predicted”.
- v. Click **Classify**, and the values under Symbol populate from 1-6 for the classes defined in our training polygons (“El\_Paso\_Training\_poly.shp”). IDs for classes are as follows: 1 for water, 2 for developed, 3 for grass and agriculture, 4 for brushland, 5 for barren land and 6 for forest. There is a 7<sup>th</sup> symbol for other (unclassified) polygons. In the table, a new field can be added to explicitly name the informational classes.
- vi. Move the Layer transparency slider to the right and click Apply with trial and error until it is transparent enough to see the imagery underneath it but not too transparent to tell what color and therefore land cover class it is.
- vii. Click OK.

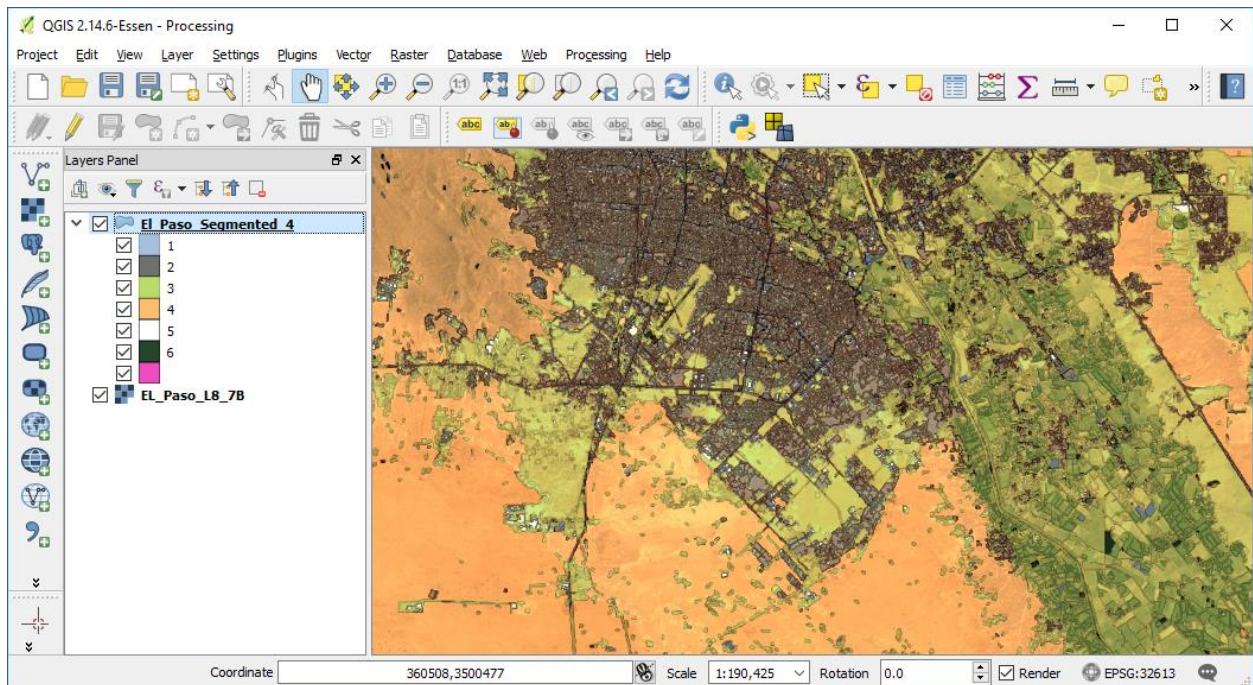


We can take now a look at the imagery.

## Instructions for Open Source Object Oriented Classification



And we can take a look at the classification on top of the imagery.




### 3.4.4 Accuracy Assessment

It can be seen that our first pass at an Object Oriented Classification is not perfect--some agricultural parcels were classified as forest. This can be helped with in-the-field ground truthing and better defined

## Instructions for Open Source Object Oriented Classification

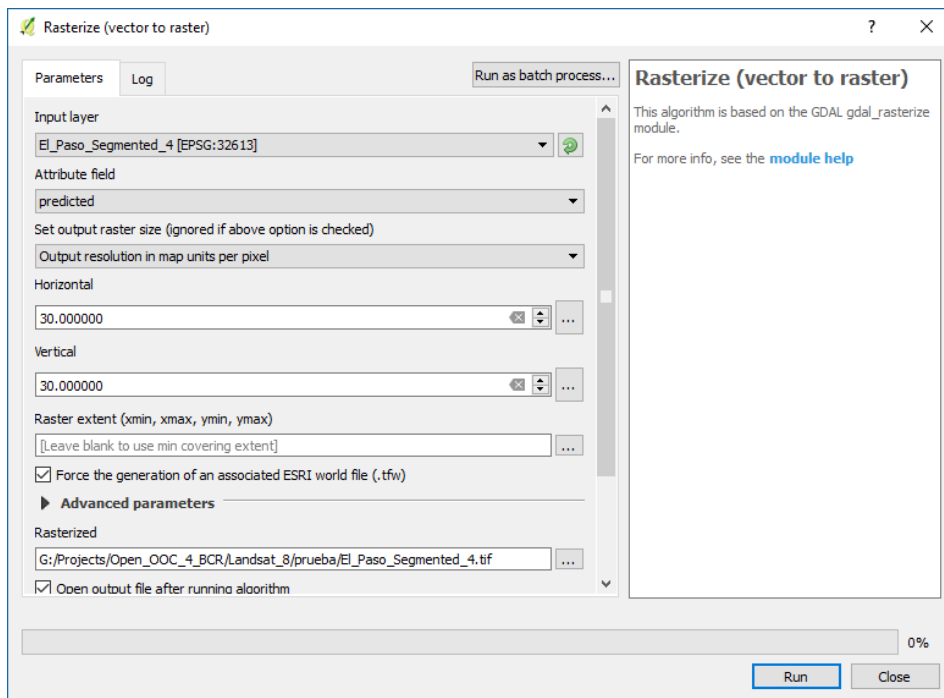
training polygons. Most classes were discernible and now, after this tutorial, you will feel adept at performing Object Oriented Classification in QGIS. Once you are visually happy with the classification results, a couple of steps may follow: dissolve contiguous polygons of the same class and perform an accuracy assessment of your classification with the second set of polygons reserved for that ("El\_Paso\_accu\_A.shp"). Information and documentation about how to perform a number of ancillary processes like how to digitize polygons, select a portion, project a shapefile, etc. can be found under the documentation tab of qgis website: <http://qgis.org/en/site/>.

To reduce the number of polygons, you may want to dissolve the polygons that are contiguous but have not been merged before, this will create a smaller shapefile that will be easier to share.

- a. You may find the dissolve tool on the top menu bar: **Vector --> Geoprocessing Tools --> Dissolve**. Uncheck the box: "Dissolve all (do not use fields)." Select the "predicted" attribute as your dissolve field, then click on  to transfer it to the Selected box. Click on the three dots by "Dissolved", select "Save to File", and name your output "El\_Paso\_Segmented\_4\_dissolved.shp".

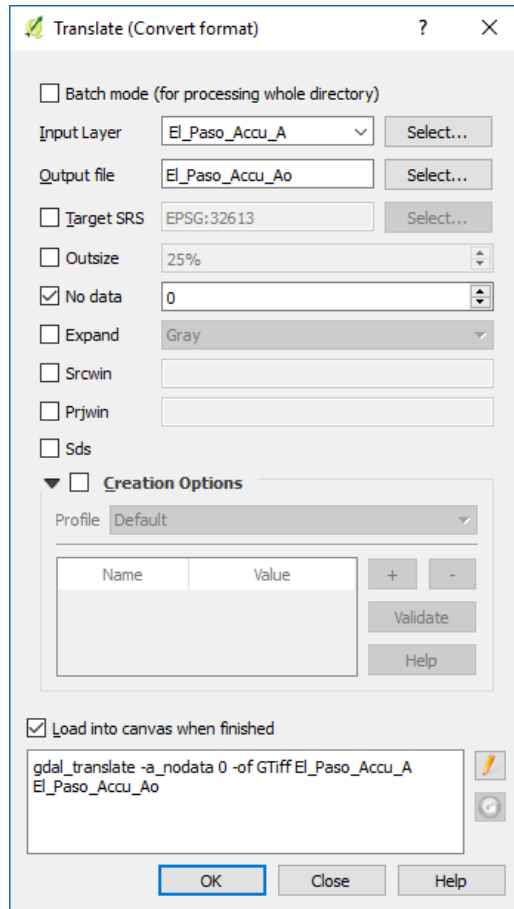
To calculate the accuracy matrix, both the referenced polygons ("El\_Paso\_accu\_A.shp", using the "Id" field"), and the resulting polygons from the classification ("El\_Paso\_Segmented\_4.shp" or "El\_Paso\_Segmented\_4\_dissolved.shp", using the "predicted" field) need to be converted into images (rasters). Accuracy assessment using completely vector products is not available at this time.

- b. On the top menu bar find: **Raster --> Conversion --> Rasterize (Vector to Raster)**
- c. Rasterize "El\_Paso\_accu\_A.shp", using the "Id" field into "El\_Paso\_accu\_A.tif", select horizontal and vertical pixel sizes about the size of the landsat pixels used for the image input, about 30 m. Click the checkmark to force the generation of an associated ESRI world file (.tfw) and click "run".



## Instructions for Open Source Object Oriented Classification

- d. Since the “El\_Paso\_accu\_A.tif”, has zeros as fillers, we don’t have a zero class, so we need to convert zeros to “No\_data”. On the top menu bar find: **Raster --> Conversion --> Translate (Convert format)**. Select “El\_Paso\_Accu\_A.tif” as your Input Layer, name your Output layer as “El\_Paso\_Accu\_Ao”, select the check mark for “No data” and assign it a “0” as value for No-data and click OK.



- e. Rasterize “El\_Paso\_Segmented\_4.shp” using the “predicted” field into “El\_Paso\_Segmented\_5.tif”, select horizontal and vertical pixel sizes about the size of the landsat pixels used for the image input, about 30 m. Click the checkmark to force the generation of an associated ESRI world file (.tfw) and click “run”. This step may take a longer time, since there is more processing.
- f. Find on the **Processing Toolbox --> GRASS GIS & commands --> Raster --> r.kappa**
- g. Select layer containing the classification result “El\_Paso\_Segmented\_4.tif” and the raster layer from the accuracy assessment polygons “El\_Paso\_accu\_A.tif”.
- h. Give a name to the error matrix, like “**El\_Paso\_Error\_Matrix.txt**”
- i. Then run **r.kappa**. This calculates the error matrix and kappa parameter for accuracy assessment of classification result.



## Instructions for Open Source Object Oriented Classification

The screenshot shows the 'r.kappa' window with the following parameters:

- Parameters** (selected tab), Log, Help
- Run as batch process...** (button)
- Raster layer containing classification result:** El\_Paso\_Segmented\_5 [EPSG:32613]
- Raster layer containing reference classes:** El\_Paso\_Accu\_Ao [EPSG:32613]
- Title for error matrix and kappa:** ACCURACY ASSESSMENT
- ☐ No header in the report
- ☐ Wide report (132 columns)
- GRASS GIS 7 region extent (xmin, xmax, ymin, ymax):** [Leave blank to use min covering extent]
- Error matrix and kappa:** G:/Projects/Open\_OOC\_4\_BCR/Deliverables/El\_Paso\_Accuracy\_Matrix.txt
- Progress bar:** 0%
- Buttons:** Run, Close

Kappa values for overall and each class are computed along with their variances. Also, percent of commission and omission errors, total correct classified result by pixel counts, total area in pixel counts and percentage of overall correctly classified pixels are tabulated. What does this all mean? The Kappa statistic (or value) is a metric that compares an Observed Accuracy with an Expected Accuracy (random chance). According to Landis and Koch: kappa values of 0-0.20 can be considered as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1 as almost perfect, meanwhile other authors like Fleiss consider kappa values > 0.75 as excellent, 0.40-0.75 as fair to good, and < 0.40 as poor.

Classification error occurs when an object (or pixel in the case of supervised or unsupervised classifications) belonging to one class (i.e. land cover) is assigned to another class. Errors of omission occur when an object is left out of the class being evaluated; errors of commission occur when an object is incorrectly included in the class being evaluated. An error of omission in one category will be counted as an error in commission in another class.

Overall accuracy classifications over 70% are normally deemed acceptable, but much depends on the purpose of it, if you want to compare land cover over time to see how well land conservation practices are working, we need to minimize error, in this case a value above 85% accuracy in the classification is preferred.

## Instructions for Open Source Object Oriented Classification

The assessment report will be written to an output file which is in plain text format and named by user at prompt while running the program. This report can then be brought into Excel or Open Office for formatting if needed.

Sample of text file of the Accuracy Assessment El\_Paso\_Accuracy\_Matrix.txt:

### ACCURACY ASSESSMENT

LOCATION: temp\_location

Mon May 08 22:08:05 2017

MASK: none

MAPS: MAP1 = (untitled) (tmp1494299281396 in PERMANENT)

MAP2 = (untitled) (tmp1494299281395 in PERMANENT)

#### Error Matrix

Panel #1 of 2

		MAP1				
	cat#	1	2	3	4	5
M	1	144	147	50	0	0
A	2	531	2709	0	0	31
P	3	7	245	6136	5060	44
2	4	0	0	0	35306	15700
	5	0	2	0	0	244
	6	0	0	175	0	0
Col Sum		682	3103	6361	40366	16019

Panel #2 of 2

		MAP1	
	cat#	6	Row Sum
M	1	0	341
A	2	0	3271
P	3	57	11549
2	4	476	51482
	5	0	246
	6	16	191
Col Sum		549	67080

Cats	% Commission	% Omission	Estimated Kappa
1	57.771261	78.885630	0.416353
2	17.181290	12.697390	0.819854
3	46.869859	3.537180	0.482200
4	31.420691	12.535302	0.211013
5	0.813008	98.476809	0.989319
6	91.623037	97.085610	0.076209

## Instructions for Open Source Object Oriented Classification

**Kappa**            **Kappa Variance**  
**0.352544**    **0.000013**

Obs	Correct	Total Obs	% Observed Correct
	44555	67080	66.420692

### **MAP1 Category Description**

1:    **Water**  
2:    **Developed**  
3:    **Grass-Agro**  
4:    **Brushland**  
5:    **Barren land**  
6:    **Forest**

### **MAP2 Category Description**

1:    **Water**  
2:    **Developed**  
3:    **Grass-Agro**  
4:    **Brushland**  
5:    **Barren land**  
6:    **Forest**

As you can see with the synthetic data used in this tutorial for training polygons and accuracy assessment polygons, the Kappa value calculated overall is .35 and the Classification Accuracy is 66%. Both values are a little under what we would like as a minimum. With field-verified data from around the time of the image collection, your kappa values and classification accuracy will be better than the ones shown in this tutorial. Hopefully that conveys the point across about the importance of field verified ground-truthing.

If the results of your classification are not what you want, you can omit training polygons and accuracy polygons with low certainty or with mixed classes. Additional measures can be to adjust parameters as the spatial radius (pixels) and range radius (radiometric) to smaller values to get smaller segments (objects) and facilitate their classification.

For additional help or information on classifications, OTB or QGIS, please keep the online help as reference:

- <https://www.orfeo-toolbox.org/documentation/>
- <https://www.qgis.org/en/site/forusers/index.html>

Additional help may be provided by contacting Miguel Pavon at [pavonma@hotmail.com](mailto:pavonma@hotmail.com) or (512) 466-3936.